

Notes

Daring the Courts: Trial and Bargaining Consequences of Minimum Penalties*

The renaissance of legislation establishing minimum penalties and other all-or-nothing sanctions¹ presents a critical question: will these

* I wish to acknowledge the many helpful comments of Bruce Ackerman, Albert Alschuler, Victoria Cashman, Zolton Ferency, Julius G. Getman, Edward Gramlich, William J. Haley, Donald Hall, Jan Kmenta, Richard Lempert, Barry Mahoney, Eli Noam, Daniel Saks, and Ed Zedlewski. I am also grateful to the National Institute of Law Enforcement and Criminal Justice, U.S. Department of Justice, for research support (Grant Award No. 78-NI-AX-0083); the views expressed in this Note, however, are not those of the National Institute.

1. The term "renaissance" significantly understates the number of recent reincarnations of the minimum-penalty concept. *See, e.g.*, S. 1437, 95th Cong., 1st Sess. § 2101(b)(1) (1977) (imposing minimum and maximum terms for felony probation); U.S. Dep't of Justice, Press Release (March 23, 1980) (legislation setting mandatory sentences was adopted in 27 states and was under consideration in 14 others during preceding three-year period). *See also* Radzinowicz & Hood, *Judicial Discretion and Sentencing Standards: Victorian Attempts to Solve a Perennial Problem*, 127 U. PA. L. REV. 1288 (1979) (describing English experience with restricted discretion and harsh penalties).

Policies restricting judicial discretion have alternated over time between legislative favor and disfavor. For example, in 1786, Tennessee eliminated capital punishment for horse theft on the first offense, noting that extensive nullification was occurring because "the punishment in its nature and gradation bear[s] no proportion to the guilt." 1 E. SCOTT, *LAWS OF THE STATE OF TENNESSEE, INCLUDING THOSE OF NORTH CAROLINA NOW IN FORCE IN THIS STATE FROM THE YEAR 1715 TO THE YEAR 1820, INCLUSIVE* 366 (Knoxville 1821) (compilation of state laws). The new statute prescribed a penalty that included a public whipping of 39 lashes, nailing of the thief's ears to the pillory and then cutting them off, and branding the thief's cheeks with the letters H and T to mark the person a horse thief. *Id.* The success of this more "lenient" penalty is unclear, for the death penalty was reinstated in 1799, *id.* at 648, but eliminated again in 1807, *id.* at 1056. This last revision allowed greater exercise of discretion than previous statutes in that it provided for, among other things, imprisonment of six months to two years, and "a number of lashes, not exceeding thirty-nine." *Id.*

The term "minimum penalty" is used in this Note to refer to any prescribed minimum sanction, including fines and imprisonment. Other forms of minimum penalty are possible. For example, a union officer found guilty of violating provisions of the Labor-Management Reporting and Disclosure Act may be barred from running for union office for a specified time period. 29 U.S.C. § 504(a) (1976). That period would be the mandated minimum penalty.

All-or-nothing sanctions are created, in effect, by the adoption of mandatory minimum-penalty policies, *e.g.*, 42 U.S.C. § 6104(a), (b) (1976 & Supp. III 1979) (penalty option under Age Discrimination Act of 1975 permitting only total cutoff of federal funds), but they also exist independently of such policies, *e.g.*, 8 U.S.C. § 1251 (1976 & Supp. III 1979) (deportation); Sup. Cr. R. 8 (disbarment). In addition, the stigma associated with criminal conviction may be considered a form of all-or-nothing penalty. *See* H. PACKER, *THE LIMITS OF THE CRIMINAL SANCTION* 36 (1968). The identifying characteristic of these penalties is the inability to divide the penalty so that a less severe level of the same form of penalty can be assigned. *Cf. Li v. Yellow Cab Co.*, 13 Cal. 3d 804, 810, 532 P.2d 1226, 1230, 119 Cal. Rptr. 858, 862 (1975) (contributory negligence doctrine is "all-or-nothing" approach).

policies work, and if not, why not?² Adopted in pursuance of a commitment to more stringent enforcement of the laws, these legislative choices of the level and form of sanction³ also restrict the discretion of various decisionmakers⁴ in the legal process. The difficulty, or paradox, of such policies is that their success depends critically on how the newly restricted discretion is exercised by those same decisionmakers.⁵ Judicial decisionmakers sometimes respond by "nullifying"

2. Whether these policies "work" is highly sensitive to the definition of success employed. This Note evaluates the success of the policies principally in terms of four considerations: (1) How well does the policy succeed on its own terms; that is, does the imposition of minimum or all-or-nothing penalties change the outcome of cases in the desired fashion? (2) How well does the policy contribute to the achievement of the policy goals more broadly defined, such as deterrence of a proscribed activity? (3) What is the incidence or pattern of eventual effect across groups and individuals? (4) How does the introduction of the policy change the process, as well as the outcome, of decisionmaking? These questions are traditional ones used in budget evaluation. See, e.g., R. MUSGRAVE & P. MUSGRAVE, *PUBLIC FINANCE IN THEORY AND PRACTICE* 179-84 (1973); Carlson, *The Status and Next Steps for Planning, Programming and Budgeting*, in *PUBLIC EXPENDITURES AND POLICY ANALYSIS* 367, 373-74 (R. Havemann & J. Margolis eds. 1970). But cf. Wildavsky, *Rescuing Policy Analysis from PPBS*, in *PUBLIC EXPENDITURES AND POLICY ANALYSIS* 461 (R. Havemann & J. Margolis eds. 1970) (criticizing program budgeting approach as administratively infeasible).

3. The concern with enforcement is not limited to the enforcement of criminal laws. See, e.g., Breit & Elzinga, *Antitrust Enforcement and Economic Efficiency: The Uneasy Case for Treble Damages*, 17 J.L. & ECON. 329, 344, 347, 352 (1974) (discussing "defects" of treble damage private remedy and recommending replacement by mandatory fine and injunctive relief available to private parties).

4. See K. DAVIS, *DISCRETIONARY JUSTICE* 9-12 (1971) (noting "variety of discretionary justice").

Policy choices that restrict judicial discretion need not be made by the legislature; they may also be made through rulemaking power legislatively delegated to the executive or judicial branches. At least in the arena of federal criminal law, however, the determination of policy must be predominantly legislative rather than judicial. *United States v. Hudson & Goodwin*, 11 U.S. (7 Cranch) 32, 34 (1812). That such a requirement is itself a restriction on judicial discretion is not coincidental. See L. FRIEDMAN, *A HISTORY OF AMERICAN LAW* 254-55 (1973) (giving examples of movements for codification in order to control judicial discretion).

5. The issue of decisionmaking under conditions of restricted outcomes is a general problem in the law. Within antitrust law, for example, the question arises in the contexts of forced full-line purchasing, tying and reciprocity arrangements, and other relationships characterized by restrictions on purchase or sale. See, e.g., Adams & Yellen, *Commodity Bundling and the Burden of Monopoly*, 90 Q.J. ECON. 475 (1976) (discussing restrictions imposed by commodity bundling); Bowman, *Tying Arrangements and the Leverage Problem*, 67 YALE L.J. 19 (1957) (discussing restrictions imposed by tying arrangements); Burstein, *The Economics of Tie-in Sales*, 42 REV. ECON. & STATISTICS 68 (1960) (discussing restrictions imposed by tie-in sales); Stigler, *United States v. Loew's Inc.: A Note on Block-Booking*, 1963 SUP. CT. REV. 152 (discussing block-booking as device for discrimination among groups of customers); cf. W. BAUMOL, *ECONOMIC THEORY AND OPERATIONS ANALYSIS* 549-69 (4th ed. 1977) (adopting activity analysis and integer programming approach).

The issue also arises in tort law. For example, the existence of a contributory negligence rule restricts the decisions of a judge or jury that might otherwise apportion damages according to relative fault of the parties. The decisionmaker may nullify by failing to find contributory negligence when it is present, by allocating damages more or less in ac-

Minimum Penalties

the law: they refuse to reach the legally correct determination of liability or guilt,⁶ thus frustrating policy intent.⁷ Further, because of the nullification possibility, pretrial negotiations, in which the prospective trial outcome serves as a benchmark, may be resolved differently with the restrictions than without them.⁸

This Note investigates the consequences of the restrictions on judicial discretion effected by the introduction of minimum or all-or-nothing penalties.⁹ A model is constructed depicting decisionmaking

cordance with comparative-negligence principles, or by changing the law through interpretation or judicial review. *See, e.g.,* *Li v. Yellow Cab Co.*, 13 Cal. 3d 804, 532 P.2d 1226, 119 Cal. Rptr. 858 (1975) (adopting comparative negligence rule after repeated failure of legislature to enact one); Rubsamen, *Medical Malpractice*, SCIENTIFIC AM., Aug. 1976, at 18-23 (examples of nullification through awarding of damages in medical malpractice cases); *cf.* Bickel & Wellington, *Legislative Purpose and the Judicial Process: The Lincoln Mills Case*, 71 HARV. L. REV. 1, 24-25, 28-35, 38, 39 & n.122 (1957) (discussing role of judicial decisions in shaping, nullifying, or standing apart from development of statutes and common law); G. Calabresi, *The Common Law Function in the Age of Statutes* 30-93 (1980) (unpublished manuscript, Yale Law School) (on file with *Yale Law Journal*) (discussing judicial interpretation of law as vehicle for change in law or nullification).

Nullification may also result from decisions by nonjudicial actors. *See, e.g.,* *Kentucky v. Dennison*, 65 U.S. (24 How.) 66 (1860) (upholding refusal of governor of Ohio to extradite person convicted of assisting slave to escape, by denying mandamus writ requested by Kentucky).

6. *See* H. KALVEN & H. ZEISEL, *THE AMERICAN JURY* 68 (1966) (finding jury nullification—as measured by jury verdict different from decision thought correct by judge—in 28% and greater leniency in 24% of cases studied); Riedel, *Discrimination in the Imposition of the Death Penalty: A Comparison of the Characteristics of Offenders Sentenced Pre-Furman and Post-Furman*, 49 TEMP. L.Q. 261, 265 (1976) (in Pennsylvania, sentences commuted more frequently when state had mandatory death penalty, and white offenders' sentences commuted more frequently than those of nonwhites); Vidmar, *Effects of Decision Alternatives on the Verdicts and Social Perceptions of Simulated Jurors*, 22 J. PERSONALITY & SOC. PSYCH. 211, 216-17 (1972) (extent of nullification by experimental subjects found to vary with set of alternative verdicts).

7. Nullification in a particular case, rather than imposition of a prescribed sanction, is an instance of policy frustration. Whether frequent and widespread nullification frustrates the overall policy is a more complicated question, turning both on the specific goals of the policy and on the effects of case-by-case nullification on other factors contributing to policy success. If, for example, the policy goal is an increase in average sentence length, then a fall in the likelihood of sanction due to frequent nullification could more than offset any increase in average penalty. *See* pp. 627-29 & notes 87-90 *infra*.

8. For discussion of the relationship between the trial and the negotiated outcomes in legal disputes, *see* D. NEWMAN, *CONVICTION: THE DETERMINATION OF GUILT OR INNOCENCE WITHOUT TRIAL* (1966), Landes, *An Economic Analysis of the Courts*, 14 J.L. & ECON. 61 (1971), and J. Lachman, *An Economic Model of Plea Bargaining in the Criminal Court System* (1975) (Ph.D. dissertation in Michigan State University Library, available from University Microfilms, Ann Arbor, Michigan).

9. There has been relatively little theoretical or empirical writing on the effects of restrictions of judicial discretion on outcomes obtained by pretrial settlement. *See* note 6 *supra*. Surprisingly, there is also a paucity of work in the industrial relations literature concerning how bargained outcomes, and negotiated versus conflict patterns of resolution, are affected by restrictions on the set of acceptable outcomes. There is in that literature, however, a concept of "prominent points of settlement," which are likely to be discussed as possible settlement outcomes. In a collective bargaining dispute, one example of a prominent point of settlement is last year's wage plus a cost-of-living adjustment. *See*

by a prosecutor or enforcement agency; the model is then used to explore the consequences of restricting discretion for individual cases and for the agency's portfolio of cases.¹⁰ The Note analyzes the effects on trial and bargained outcomes of various levels of evidence and legally irrelevant information, of alternative levels of agency enforcement resources, and of the severity of the minimum or all-or-nothing penalty. After considering the conditions under which policy intent is effectively implemented, the Note concludes that the success of such legislation is far from certain. Policies restricting judicial discretion may backfire, resulting in lower levels of sanction actually being assigned and less deterrence of the activity proscribed. The policies most likely to succeed are those that take the prospective nullification response into account in the initial determination of the form and severity of sanction.

I. Judicial Nullification in the Face of Restricted Outcomes

The success of legislation creating minimum or all-or-nothing penalties is endangered by the prospect of judge or jury nullification at trial—the refusal or diminished readiness to reach the legally correct determination of liability or guilt.¹¹ Nullification by judicial decisionmakers¹² that is precipitated by restrictions on outcome follows a tradition extending from as far back as medieval England¹³ through

R. WALTON & R. MCKERSIE, A BEHAVIORAL THEORY OF LABOR NEGOTIATIONS 41-45, 51-52, 105-06 (1965) (prominent points are more likely than others to be considered as possible settlements, and it is good bargaining strategy to make desired settlement points appear as prominent ones to other party).

10. The text depicts decisionmaking from the perspective of an enforcement agency or prosecutor. The problem as viewed by a defendant is analogous, with the defense's objective being stated in terms of obtaining a desirable case outcome and a high level of consumption, or whatever characterization is chosen for the other goods and activities entering into the preference function.

11. See M. KADISH & S. KADISH, DISCRETION TO DISOBEY 47, 55-56 (1973); Schefflin, *Jury Nullification: The Right to Say No*, 45 S. CAL. L. REV. 168, 168, 173, 181-85 (1972).

12. Defining nullification is problematic in itself, for to say that a decisionmaker has nullified the law presupposes that what is required for fulfillment of the law is known with sufficient precision to exclude the decision taken. If this definitional requirement is rigidly imposed, every reversal on appeal might be construed to represent an appellate determination that the trial judge has nullified; indeed, appellate review has been used for the purposes of correcting and disciplining trial judges and restricting their discretion. See, e.g., *United States v. Lynd*, 349 F.2d 785, 787-90 (5th Cir. 1965) (appellate court directs trial court in civil rights voting case to issue appellate court's injunctive order). But cf. Jeffries & Stephan, *Defenses, Presumptions, and Burdens of Proof in the Criminal Law*, 88 YALE L.J. 1325, 1393-95 (1979) (discussing case concerning constitutionality of kidnapping statute, and concluding that trial court decision was correct, but appellate court's result was "bizarre").

13. See *Sykes v. Director of Pub. Prosecutions*, [1962] A.C. 528, 554-55 (Lord Denning)

Minimum Penalties

twentieth-century America.¹⁴ Put simply, a major problem with minimum and all-or-nothing penalties is that they may not be applied at all.¹⁵ Although such nullification need not be considered a failure of judicial function,¹⁶ it may nevertheless signal a failure to fulfill legislative intent.¹⁷

If the new policy is intended to change the level of penalty, but is not directed either at changing the set of actions to which liability attaches or at decreasing the likelihood of sanction, then frequent

(as early as 1484, when defendants were charged with capital offense of treason, judges sometimes gave instruction on misprision as well); *cf.* *Bushell's Case*, 22 Charles II [Vaughan's Reports 135] 6 St. Tr. 999, 124 Eng. Rep. 1006 (Ex. 1670) (court's imprisonment of juror for refusal to convict defendant found to be impermissible).

14. Judge and jury nullification has been noticed especially in times of social upheaval in the United States, such as the period of the fugitive slave laws in the mid-nineteenth century and the era of the Vietnam War. *See, e.g.,* R. COVER, *JUSTICE ACCUSED* 191, 211-14 (1975) (discussing controversies surrounding jury discretion and widespread acquittals in fugitive slave cases); Kunstler, *Jury Nullification in Conscience Cases*, 10 VA. J. INT'L L. 71 (1969) (discussing jury nullification in Vietnam-era dissent cases).

15. *See* W. FORSYTH, *HISTORY OF TRIAL BY JURY* 430-34 (London 1852) (convictions infrequent under regime of harsh penalties).

16. The desirability and functions of jury nullification have been the subject of considerable controversy. Nullification may be viewed from several different perspectives. It may be seen as the abuse of discretion and public trust; as an expression of public preferences and, as such, a part of the regenerative processes of law; as the fulfillment of jurors' commitment to fairness; as either the responsible or the irresponsible exercise of judicial discretion, depending on the reasons for nullifying and on the fairness of the resulting pattern of gainers and losers among defendants; or as the constructive use of inconsistency, satisfying simultaneous public desires for a statement of moral principles in the abstract and for individualized justice. *See* W. FORSYTH, *supra* note 15, at 431-32 ("Juries would not condemn men to the gallows for an offense [forgery of bank-notes] of which the punishment was out of all proportion to the crime . . ."; such nullification led to "amending many bad laws which the judges would have administered with professional bigotry.") (quoting Lord John Russell); M. KADISH & S. KADISH, *supra* note 11, at 45-66, 139-40 (nullification by judicial decisionmakers and use of prosecutorial discretion is "legitimated interposition" for achievement of social ends); Schulhofer, *Harm and Punishment: A Critique of Emphasis on the Results of Conduct in the Criminal Law*, 122 U. PA. L. REV. 1497, 1554-57 (1974) (discussing sources of decisionmaker dissatisfaction giving rise to nullification, and possibility of disparate impact as function of defendant characteristics such as race).

17. *See, e.g.,* Richardson & Heilweil, *Sentencing Patterns Under the 1973 New York State Drug Laws*, in *STAFF WORKING PAPERS OF THE DRUG LAW EVALUATION PROJECT* 245, 259 (Joint Committee on New York Drug Law Evaluation ed. 1978) (despite marked increase in percentage of Class A drug convictions resulting in prison sentences, actual number of convictions for which prison sentences were imposed fell from 2,039 in 1972 to 1,164 in 1975); *cf.* Sajovic & Goldsmith, *The Effects of the 1973 Drug Laws on Heroin Use in New York City*, in *STAFF WORKING PAPERS OF THE DRUG LAW EVALUATION PROJECT* 1, 8 (Joint Committee on New York Drug Law Evaluation ed. 1978) (despite introduction of stiff minimum penalties for drug use, narcotics use in New York City, which had been declining prior to new laws, became stable). *But cf.* L. FRIEDMAN, *supra* note 4, at 253 (uncertainty about consistency in applying laws, resulting from "[t]he degeneracy (or overvigilance) of judge, jury, state, or national authority," led to periodic lynching and vigilante outbursts as popular response to perceived judicial nullification).

nullification undermines the overall policy.¹⁸ When discretion over the level of sanction is foreclosed,¹⁹ the liability or guilt decision be-

18. In addition to the problems of policy failure in the senses described above, see notes 2 & 16 *supra*, there may be far-reaching though less easily observed ramifications. See, e.g., H. PACKER, *supra* note 1, at 262-68, 286-89 (laws frequently nullified in courts or in earlier enforcement processes create environment in which criminal sanction is diluted, increasing likelihood that other laws will be disobeyed); Alschuler, *Sentencing Reform and Prosecutorial Power: A Critique of Recent Proposals for "Fixed" and "Presumptive" Sentencing*, 126 U. PA. L. REV. 550, 551 (1978) (fixed or presumptive sentences "concentrat[e]" discretion in such a way that benefits are unfairly distributed).

Anticipated trial and settlement outcomes indicate to prospective litigants the institutional consequences of their behavior, and thus affect behavior in cases that do not come to the attention of a court. By this form of long-distance decisionmaking, the message from patterns of trial and settlement outcomes may clearly indicate how a dispute would be resolved in the legal process and thus save the potential litigants the trouble and expense of going to the court to find out. For example, if cases with similar facts involving alleged discriminatory discharge of employees for union activity or organizing efforts always yield similar results, one might expect that, in the future, similar cases would be informally resolved by the disputants and that the National Labor Relations Board would receive fewer complaints than the number of disputes it effectively settles. A second message to potential litigants, however, may be that the costs associated with obtaining a trial or pretrial settlement are substantial; if so, prospective litigants whose dispute involves low anticipated benefit may choose not to undertake such litigation.

To the extent that the first message is clear, the courts and administrative agencies may be credited with doing a good job of "making law," assuming that the clear decision is a desirable one. The second kind of message is more problematic. Because decisions not to enter the legal process are made in an atomized fashion, courts and policymakers are unaware of the frequency and reasons for this behavior, and thus may misunderstand not only the "discouraged litigant" problem but also the operation of the more visible processes of dispute resolution that the decision of the discouraged litigant reflects. Cf. A. SPENCE, MARKET SIGNALING 1-30, 152-76 (1974) (describing information, in addition to that of traditional economic variables, relevant to labor market decision-making).

In addition, the fulfillment of public purposes may depend largely on private actions through litigation, particularly in areas where private rights of action are substitutes for public enforcement. See, e.g., Calabresi, *The Problem of Malpractice—Trying to Round Out the Circle*, in THE ECONOMICS OF MEDICAL MALPRACTICE 233, 233-38 (S. Rottenberg ed. 1978) (discussing incentives arising from tort law, from regulatory or criminal law, and from contract law for avoidance of costs of medical accidents); Shavell, *Theoretical Issues in Medical Malpractice*, in THE ECONOMICS OF MEDICAL MALPRACTICE 35, 50-55 (S. Rottenberg ed. 1978) (comparing alternative approaches to liability for medical accidents in terms of their effects on quality of medical care). As a result, disputes not worth litigating because of the high costs of the legal process may be resolved poorly, with a consequent social loss. Finally, the resulting distribution of gainers and losers in the legal process may change in politically unacceptable ways as a function of the costs associated with legal dispute resolution. See, e.g., Note, *Costs and the Plea Bargaining Process: Reducing the Price of Justice to the Nonindigent Defendant*, 89 YALE L.J. 333, 351-52 (1979) (nonindigent defendant, who must actually bear costs of trial even if acquitted, is subject to special pressures to plead guilty).

19. The Vidmar experiments on jury decisions under varying choice constellations and the Kalven and Zeisel findings point out more explicitly the danger of using the liability or guilt decision to express sanction preferences. See H. KALVEN & H. ZEISEL, *supra* note 6, at 68 (jury outcomes sometimes more lenient than judge decisions even when both would convict); Vidmar, *supra* note 6, at 216 (experimental juries acquitted more frequently when their discretion was restricted).

Nullification occurs, when it otherwise would not, because of the presence of the re-

Minimum Penalties

comes the major vehicle for nullification.²⁰ More specifically, if the decision on liability is in effect a yes or no decision regarding the required penalty, the decisionmaker is faced with assigning the mandated penalty or assigning no penalty at all.²¹

striction on outcome. However, the occurrence of nullification will be associated with differences among cases and individual parties rather than with the presence of the penalty provisions. For example, the decisionmaker will explain that the penalty was considered inappropriate or unfair because of factors such as weakness of evidence, lack of harmful consequences, or characteristics of or personal hardship to the defendant. Kalven and Zeisel found that evidence factors were said to be the basis for disagreement in 78% of the cases in which the jury would have been more lenient than the judge, with this figure rising to 84% for hung juries; for cases in which the jury would have been more harsh, convicting when the judge would not, evidence factors were cited 93% of the time. H. KALVEN & H. ZEISEL, *supra* note 6, at 111; cf. P. DEVLIN, *THE ENFORCEMENT OF MORALS* 21 (1959) ("I do not mean that they often deliberately disregard the law. But if they think it is too stringent, they sometimes take a very merciful view of the facts.")

With respect to the importance of harmful consequences, see G. FLETCHER, *RETHINKING CRIMINAL LAW* 472-83 (1978) (discussing harmful consequences as giving rise to pattern of criminal liability), and Schulhofer, *supra* note 16, at 1554-57 (lack of harmful consequences may increase likelihood of nullification by judicial decisionmakers); for the effects of defendant characteristics, see H. KALVEN & H. ZEISEL, *supra* note 6, at 105 (various characteristics of defendant are associated with more frequent nullification by juries), and Schulhofer, *supra* note 16, at 1527-30, 1559-63 (discussing effects of defendant characteristics on guilt determination).

20. Juries and judges have found other vehicles for expressing dissatisfaction with restrictions placed on their decisionmaking. See, e.g., *Commonwealth v. Franklin*, 172 Pa. Super. Ct. 152, 152, 157-61, 92 A.2d 272, 273, 275-77 (1952) (despite acquittal, defendant ordered held on bail for two years after trial, in keeping with Pennsylvania tradition dating from at least 1799); *Braynes case*, discussed in *Bushell's Case*, 22 Charles II [Vaughan's Reports 135, 153] 6 St. Tr. 999, 124 Eng. 1006, 1015 (Ex. 1670) (jury attempted "to delude the Court, by agreeing upon two verdicts, and concealing the latter, if the Court would be satisfied with the former"); Newman, *To Rid the U.S. Courts of the Slovik Syndrome*, N.Y. Times, Sept. 21, 1979, § A, at 26, col. 4 (having "misplaced confidence" that subsequent decisionmakers will mitigate sanctions ordered, judges sometimes feel liberated from having to make decision they would be satisfied with if actually carried out). For a model of juror decisionmaking under the assumption that jurors have single-peaked preferences, see B. Grofman, *Mathematical Models of Juror and Jury Decisionmaking: The State of the Art* 15-24 (1977) (unpublished manuscript, University of California, Irvine) (on file with *Yale Law Journal*). See generally Davis, Bray, & Holt, *The Empirical Study of Decision Processes in Juries*, in *LAW, JUSTICE, AND THE INDIVIDUAL IN SOCIETY* 326 (J. Tapp & F. Levine eds. 1977).

21. In *Beck v. Alabama*, 447 U.S. 625 (1980), the Supreme Court struck down an Alabama statute that allowed juries to choose only between a death penalty conviction and acquittal, without a choice of any lesser included noncapital offense, because the restricted choice might result in juries convicting or acquitting for impermissible reasons. By contrast, in *Rummel v. Estelle*, 445 U.S. 263 (1980), the Court upheld a Texas statute providing for mandatory life imprisonment upon a third felony conviction, as applied to a defendant convicted for obtaining \$120 under false pretenses; it defined the reach of the holding by noting "the unique nature of the death penalty for purposes of Eighth Amendment analysis." *Id.* at 272.

To the extent that nullification does occur, those left behind after the judicial "filter" is applied can be characterized as the individuals "chosen" to receive punishment. See C. BLACK, *CAPITAL PUNISHMENT: THE INEVITABILITY OF CAPRICE AND MISTAKE* 21 (1974); cf. Riedel, *supra* note 6, at 275-83 (statistical profile of death row inmates showing changes in demographic and evidentiary characteristics of cases after introduction of mandatory

Prescribing a binding minimum penalty for situations in which such a restriction previously did not exist effectively introduces indivisibility in the units of outcome. This situation is analogous to that faced by individuals making economic decisions about the purchase of automobiles, houses, or packs of soft drinks, in which the option of a fractional quantity is not available.²² When the presence of a minimum penalty renders unattainable the outcome otherwise desired, the judicial decisionmaker must decide whether the next best solution is to accept the minimum or to nullify.²³ If decisionmakers

death penalties); Wolfgang & Riedel, *Race, Judicial Discretion and the Death Penalty*, 407 ANNALS 119 (1973) (same).

Viewed from this perspective, the findings of the Chicago jury project are consistent with the contention that differences in outcome will be attributed to differences among individuals or evidence, rather than to the level of the penalty itself. Although some of the judge-jury disagreement was attributable to concern with the harshness of the penalty, most was attributable to dissatisfaction with "the evidence" or to "values" differences between judge and jury. Kalven and Zeisel found that "facts" by themselves were a cause of disagreement in 34% of the cases in which judge and jury disagreed; that "values alone," such as defendant characteristics, were a cause in 21% of the cases; and that values and facts together were a cause in the remaining 45% of the cases. H. KALVEN & H. ZEISEL, *supra* note 6, at 116, 165-67. Their results indicate that various factors may be heavily intermingled with one another.

[T]he jury does not often consciously and explicitly yield to sentiment in the teeth of the law. Rather it yields to sentiment in the apparent process of resolving doubts as to evidence. The jury, therefore, is able to conduct its revolt from the law within the etiquette of resolving issues of fact.

Id. at 165.

In many circumstances, the jury would prefer to make no choice at all; they would prefer to leave the choice in the hands of the judge. [1969] 8 PERSONAL INJURY VALUATION HANDBOOK (Jury Verdict Research) 5335, 5335. This interpretation could also apply to the 5.5% hung juries found in the Kalven and Zeisel study. See H. KALVEN & H. ZEISEL, *supra* note 6, at 56.

22. In the field of marketing research, virtually no information is available about the process of making purchasing and packaging decisions. First, what studies there may be have for the most part been made by businesses that have a strong incentive to keep the results secret. Second, studies that have asked consumers how they reached their decisions and compared the subjective results with purchasing patterns suggest some inconsistency in the subjective and "objective" rankings of factors by importance. See, e.g., Permut, *Cue Utilization Patterns in Student-Faculty Evaluation*, 83 J. PSYCH. 41, 46-47 (1973). Finally, there appears to be no general theory of how purchasers determine the purchase-no purchase threshold as a function of packaging of the product, nor any theory of the optimal package size as a function of anticipated purchasing decisions. But cf. J. HIRSHLEIFER, INVESTMENT, INTEREST, AND CAPITAL 66-69, 108 (1970) (economic analysis of capital goods purchases in discrete units).

Related to the question of decisionmaking under conditions of indivisibility is the theory of the second best, which is more accurately a theory of what the second best is not. When the optimal outcome in a situation cannot be attained, the next best outcome may not be the outcome remaining that is closest to the original optimum. See J. HENDERSON & R. QUANDT, MICROECONOMIC THEORY 286-288 (2d ed. 1971) (discussion of second best theory in welfare economics context).

23. Juries may have the responsibility of imposing sanctions, but more typically they are required only to make a determination of liability or guilt. If they are not told the

Minimum Penalties

respond by assigning the new penalty in cases in which they would otherwise assign a lesser one, the intent of the policy is fulfilled. However, if decisionmakers respond by failing to find liability altogether when they otherwise would,²⁴ the policy is frustrated. The extent of judicial nullification may, of course, be sensitive to the severity of the all-or-nothing or minimum penalty.²⁵

consequences of their liability determination, they nevertheless make the decision based on whatever information about the consequences they do have, even though it may be sketchy or incorrect. *See, e.g.,* [1969] 8 PERSONAL INJURY VALUATION HANDBOOK, *supra* note 21, at 5335 (42% of jurors assumed defendant was covered by insurance in some or all cases, and said their decision would not have been different had they been sure); Kalven, *The Jury, the Law, and the Personal Injury Damage Award*, 19 OHIO ST. L.J. 158, 165-69 (1958) (jurors use information going to issue of damages in determining liability).

24. *See, e.g.,* S. POISSON, RECHERCHES SUR LA PROBABILITÉ DES JUGEMENTS 390-91 (Paris 1837) (when, in 1832, French juries were given additional verdict option of finding guilt under extenuating circumstances, conviction rate rose by 10%) (percentage compiled from author's numbers). Vidmar arrived at similar conclusions for experimental juries. He hypothesized that under conditions of restricted decision alternatives, the higher the least severe guilt alternative, the greater the chances of obtaining a not-guilty verdict. Vidmar, *supra* note 6, at 215. When the jurors were allowed to choose only between first-degree murder and acquittal, they acquitted 54% of the time, but when two intermediate choices were added, the acquittals accounted for only 8% of the dispositions. *Id.*; *see* Larntz, *Reanalysis of Vidmar's Data on the Effects of Decision Alternatives on Verdicts of Simulated Jurors*, 31 J. PERSONALITY & SOC. PSYCH. 123, 123-25 (1974); B. Grofman, *Theory and Experiment on Juries: Did the Supreme Court Get It Right?* (June 1977) (Social Science Working Paper No. 124, School of Social Sciences, University of California, Irvine).

25. Two mechanisms exist for monitoring the use of discretion. If juror or judge dissatisfaction with the penalty is sufficiently widespread, it may be reflected in statistical distributions of trial outcomes and be recognized as resulting from the penalty requirement. *See, e.g.,* P. GREENWOOD, S. WILDHORN, E. POGGIO, M. STRUMWASSER, & P. DELEON, PROSECUTION OF ADULT FELONY DEFENDANTS IN LOS ANGELES COUNTY 48-59 (1973) (tables depicting disposition of cases by ethnic group of defendants and by other characteristics) [hereinafter cited as LOS ANGELES FELONY PROSECUTIONS]; Richardson & Heilweil, *supra* note 17, at 291-302 (tables and discussion of association of prior record, age, quantity of heroin or type of drug with likelihood of prison sentence for defendants who pleaded guilty to drug charges). Appellate review also provides a safeguard against unacceptable decisions on a case-by-case basis. *See* Note, *Courting Reversal: The Supervisory Role of State Supreme Courts*, 87 YALE L.J. 1191 (1978).

In the pretrial context, neither of those devices necessarily signals dissatisfaction with the penalties or unacceptable patterns in their application. Problems that arise from differences in the initial strength of the parties are less easily discernible in cases resolved without trials. Alschuler contends that the potential unfairness and abuse of discretion by prosecutors in plea bargaining are usually beyond the reach of outsiders seeking to monitor the administration of justice. Alschuler, *supra* note 18, at 551. Employing what one might call a "hydraulic" theory of judicial process discretion, in which restrictions on the discretion exercised by one decisionmaker precipitate compensating adjustments by another, he argues that the introduction of fixed or presumptive sentencing concentrates discretion in "an inappropriate agency . . . in which the benefits of this discretion are made available only to defendants who sacrifice their constitutional rights." *Id.* at 551; *see* TWENTIETH CENTURY FUND TASK FORCE ON CRIMINAL SENTENCING, FAIR AND CERTAIN PUNISHMENT 11-34 (1976) (study and recommendation concerning presumptive sentencing).

II. Pretrial Settlements With and Without Restrictions on Judicial Discretion

Negotiated settlements account for up to ninety-five percent of the convictions obtained in criminal cases handled in urban courts²⁶ and for the great majority of civil cases brought in state and federal courts.²⁷ Despite the seeming rarity of dispute resolution by trial, the pattern of trial outcomes generally and the anticipated trial outcome in a given case are critical factors in determining the level of bargained outcomes. The anticipated trial outcome serves as a benchmark to the parties when they assess the desirability of possible bargains and decide whether to accept a negotiated settlement or to reject the final settlement offer in favor of going to trial.²⁸ Thus, changes in the

26. See D. McINTYRE, *LAW ENFORCEMENT IN THE METROPOLIS* 132 (1967) ("[T]he magnitude of guilty pleas [as a percentage of felony convictions], on a national scale, has been estimated to be as high as 95 percent and as low as 69 percent.") (footnotes omitted); Lachman & McLaughlin, *Models of Plea Bargaining*, in *MODELING THE CRIMINAL JUSTICE SYSTEM* 145, 147 (S. Nagel ed. 1977) (guilty plea yielded more than 16 times as many convictions as did trials in Detroit in 1972, and estimates of bargained dispositions in American cities range from 75% to 95%).

27. See H. ROSS, *SETTLED OUT OF COURT* 4-6 (1970) (settlements without trial are ubiquitous, accounting, for example, for 98% of bodily injury negligence claims in New York).

28. As with the desirability of and purposes served by nullification at trial, there are conflicting perspectives on the desirability of and purposes served by pretrial negotiations. One argument in favor of pretrial settlements is that of expediency: settlement of disputes without recourse to adversary proceedings yields an outcome that is desirable to both parties—that is, the outcome is Pareto optimal—and does so at substantially lower cost than that of a full trial. See, e.g., *Santobello v. New York*, 404 U.S. 257, 261 (1971) (suggesting that judicial process depends critically on existence of plea bargaining); D. NEWMAN, *supra* note 8, at 4 ("Compared to the typically long, costly, and complex trial, the guilty plea is a model of efficiency . . .") This perspective often implicitly assumes that, in the absence of "caseload pressure," few disputes would be resolved by pretrial bargaining and that such a pattern of resolution would be preferable to the present one. *But cf.* M. HEUMANN, *PLEA BARGAINING* 27-33 (1977) (finding plea bargaining prevalent and equally important in cities of varying sizes and varying levels of resources and caseloads). Other writers have argued that the pattern of widespread settlement of disputes reflects not a response to caseload pressure, but an institutional style of operation, which is perpetuated by the socialization of newcomers to the staffs of enforcement agencies. See, e.g., *id.* at 92-126.

The legal debate over negotiated settlements has tended to focus on the desirability of pretrial bargaining apart from considerations of cost to the parties or the possible desire of parties to avoid trial for reasons other than cost saving. One side of the debate has emphasized the importance of the right to trial and of other rights effectively waived or compromised by obviating the need for trial through negotiating a settlement. See, e.g., Alschuler, *supra* note 18; see Note, *Guilty Plea Bargaining: Compromises by Prosecutors to Secure Guilty Pleas*, 112 U. PA. L. REV. 865, 869-70, 874-78 (1964). The concern with pretrial bargaining is with its incidence—by race, income, age, seriousness of violation or harm—and with the conditions of exchange, that is, the rights sacrificed in return for penalty or charge reductions.

For a behavioral explanation in a general bargaining context, see T. SCHELLING, *THE STRATEGY OF CONFLICT* (1960), and R. WALTON & R. MCKERSIE, *supra* note 9. Economic

Minimum Penalties

anticipated trial outcome, such as the increased possibility of nullification, result in corresponding changes in negotiated settlements. These, in turn, will have important consequences for the success of such policies as minimum or all-or-nothing penalties.

A. *Pretrial Settlements Without Restrictions*

The traditional explanation of why parties bargain is that they agree to forego their right to a trial in exchange for a more favorable outcome as measured by the level of penalty.²⁹ This condition, however, may not and need not be met.³⁰ Even if the pretrial settlement were to yield the same outcome that would have resulted from trial, both parties save the costs of going to trial³¹ as well as the costs associated with delay. The consequent savings to each party provide the "play" that defines the limits of bargaining. A bargainer such as a prosecutor or enforcement agency is better off³² accepting a pretrial

models of bargaining typically assume that the benchmark effect of anticipated outcomes has already been determined and focus instead on how the bargainers divide the amount of the bargained good in the settlement range. *See, e.g.*, F. ZEUTHEN, PROBLEMS OF MONOPOLY AND ECONOMIC WARFARE 115-21 (1930) (bargaining range assumed to be given); Bishop, *Game-Theoretic Analyses of Bargaining*, 77 Q.J. ECON. 559, 563 (1963) (discussing models of bargaining under assumption that Pareto optimal set is known).

29. *See, e.g.*, Newman, *Pleading Guilty for Considerations: A Study of Bargain Justice*, 46 J. CRIM. L.C. & P.S. 780, 783-85 (1956); Note, *supra* note 28, at 874-78.

30. *Cf.* W. Rhodes, *Plea Bargaining: Who Gains? Who Loses?* (1978) (Institute for Law and Social Research, PROMIS Series No. 14) (statistical study finding that defendants pleading guilty in District of Columbia obtain sentences on average no lower than those of persons going to trial). Interestingly, this explanation is not the traditional one in the labor context. Indeed, the early theory of Sir John Hicks described the benefit of "skilled negotiation" as that of reaching the same outcome at an earlier time without the costs of an extended strike. J. HICKS, *THE THEORY OF WAGES* 143-44 (1st ed. 1932).

31. *See* United States v. Wiley, 184 F. Supp. 679, 685 (N.D. Ill. 1960) (estimating government's costs of instant trial at \$11,250 until time of sentencing); Note, *supra* note 18, at 351-52 (discussing costs of disposition by trial or plea bargaining for nonindigent defendant); J. Lachman, *supra* note 8, at 29-31, 60-62 (discussing costs associated with trial and bargain outcomes).

32. Economic models have adopted alternative characterizations of the bases for enforcement agency decisionmaking, including maximizing the expected number of convictions weighted by sentence length, *see, e.g.*, Landes, *supra* note 8, at 64, making prosecution decisions, viewed as investment, so as to reduce future criminal activity, *see, e.g.*, Forst & Brosi, *A Theoretical and Empirical Analysis of the Prosecutor*, 6 J. LEGAL STUD. 177, 180-84 (1977), and maximizing utility, defined as a function of two variables representing sentence outcomes in different cases, without assigned weights, *see, e.g.*, J. Lachman, *supra* note 8, at 28-29. The third approach—assuming the enforcement agency's or prosecutor's goal to be a combination that includes seeking favorable outcomes from individual cases, as well as disposing of all cases in the docket—is employed here in order to investigate the contention that pursuit of convictions in cases to which a mandated minimum penalty applies may meet with success at the expense of lower outcomes for other cases in the prosecutor's or enforcement agency's portfolio.

settlement up to the point that the cost savings associated with avoiding trial exactly offset the difference in desirability between the trial outcome and the pretrial settlement. This cutoff point is a bargainer's reservation settlement.³³

Although this pretrial settlement outcome may itself be less desirable than the trial outcome, the cost savings associated with trial and with delay until trial allow a prosecutor or enforcement agency³⁴ to allocate more resources to the remaining cases in its portfolio.³⁵ When this reallocation occurs, the "loss" in the instant case is associated with a "gain" in outcome for the other cases in the portfolio.

1. *The Trial Outcome Benchmark*

Before entering into bargaining, a party must determine its bargaining stance: it must decide how much it is willing to concede before invoking its right to go to trial.³⁶ Determination of this reservation settlement is made by comparing the anticipated trial outcome with successively less attractive levels of bargained outcome until the level is reached that yields exactly the same satisfaction as the anticipated trial outcome. The area between the cutoff points of the two parties constitutes the bargaining zone.³⁷

33. The "reservation-price settlement" or "reservation settlement" is that level of bargained outcome at which a party is indifferent between settlement and the prospect of going to trial. This is analogous to the definition of the reservation wage in labor economics, and is variously described by the terms "concession point," or "resistance point," or by other terms in the bargaining literature. See, e.g., R. WALTON & R. MCKERSIE, *supra* note 9, at 43.

34. An analogous structure describes the determination of the defendant's reservation price. See note 10 *supra*. However, depending on how the concept of risk aversion is represented in the context of undesirable goods such as sentences or fines, the defendant's reservation settlement function may be discontinuous, a result consistent with defendant specialization in case outcome or in nontrial goods—for example, consumption goods. See J. Lachman, *supra* note 8, at 66-79.

35. See Rhodes, *The Economics of Criminal Courts: A Theoretical and Empirical Investigation*, 5 J. LEGAL STUD. 311, 321-28 (1976) (levels of bargained settlements of criminal cases found to be positively related to level of prosecutor resources in cross-sectional analysis by city).

36. The choice of going to trial is a weapon in pretrial negotiations. See H. Ross, *supra* note 27, at 215-16. A dispute that does go to trial may be construed as a failure of negotiations. *Id.*

37. The agreement, if one is reached, should reside in this bargaining zone, as depicted in Figure A. The agreement reached by the parties will usually be approved by the judge. See *McCoy v. United States*, 363 F.2d 306, 307 (D.C. Cir. 1966) (refusal to accept guilty plea should be for "good reason"); Alschuler, *The Trial Judge's Role in Plea Bargaining, Part I*, 76 COLUM. L. REV. 1059, 1064, 1067-69, 1153-54 (1976) (discussing frequency with which judges approve parties' agreements and advocating greater judicial control of plea bargaining process); Note, *Judicial Discretion to Reject Negotiated Pleas*,

Minimum Penalties

Thus far, the discussion has assumed that the level of resources of the agency, the initial strength of the parties relative to one another, and the resulting agency resource-allocation pattern are all given. Alternative values of any of these items will affect the level of the bargaining cutoff.

Agency Resources. The agency's resource level represents only those resources freely allocable between two cases. It is measured by the difference between the agency's total budget and its fixed costs, where the latter are costs that do not vary with outcome. An increase in the agency's allocable resources—for example, because of a reduction in case-maintenance costs due to technological advance—would yield a different resource level.³⁸ As a result, more resources would be devoted to the instant case and others, thus improving the anticipated trial outcome. This more favorable trial outcome would, other things being equal, require that a bargain be more favorable to the agency in order to be as satisfying as the new anticipated trial outcome.

63 GEO. L.J. 241 (1974) (prosecutorial abuse of discretion required in order for judge to reject bargain). For purposes of the argument in the text, it is assumed that such refusal by the judge does not occur.

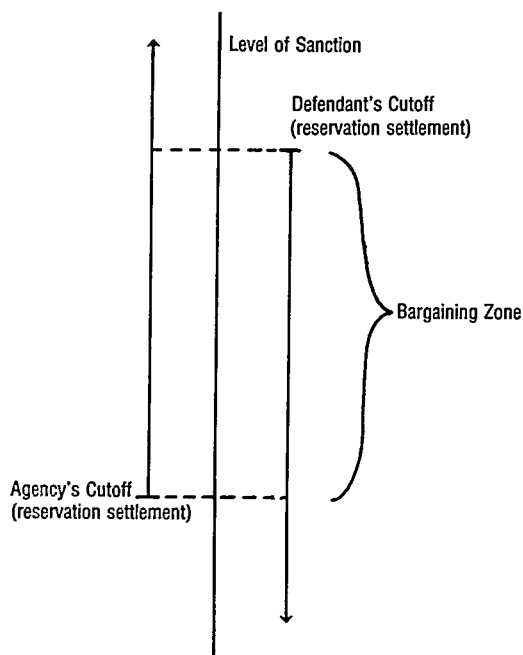


Figure A. Cutoff Points of the Parties, and Bargaining Zone.

38. See J. Lachman, *supra* note 8, at 29-31, 88-89 (discussing enforcement agency costs and resource effects); *id.* at 59-62 (discussing costs to defendant).

Relative Strength of the Parties. The initial strength of the parties is a function of specified factors giving rise to the dispute and of the characteristics of the parties. These factors and characteristics include evidence admissible at trial as well as legally irrelevant information.³⁹ For example, an admission of wrongdoing is relevant information, and the race of the defendant is irrelevant information; both affect the relative strength of the parties.⁴⁰ The weaker the initial position of the defendant relative to that of the agency, the better the anticipated trial outcome⁴¹ and the more favorable the reservation settlement level must be to the agency in order to make pretrial settlement as attractive an option as going to trial.⁴²

Agency Resource Allocation. Implicit in the determination of the anticipated trial outcome and its associated satisfaction level is a

39. The relative strength relation may be written

$$r = r(e, d_1, d_2, \dots, d_h)$$

where e denotes admissible evidence and d_i , $i = \{1, 2, \dots, h\}$, represents characteristics of the defendant, such as race, age, sex, employment status, previous relationship with the agency or other governmental units, and other factors. Alternative states of any of the contributing factors can be expected to affect the level of r , and consequently the trial outcome level. For example, if the defending party has admitted to engaging in the proscribed activity, the value of r is lower and is consequently associated with trial and bargained outcomes more favorable to the prosecutor.

Although a disadvantage in initial strength may be overcome to some extent by expenditure of resources for case development, the party so disadvantaged will, other things being equal, always be at some disadvantage. This occurs either because the initial handicap is not mitigated by sufficient expenditure of resources, or because the resources used to compensate for previous disadvantage could otherwise have been devoted to improving outcome still further. For handicaps in the form of evidentiary information, this effect may or may not be acceptable. For legally irrelevant information, it is clearly undesirable. See E. GREEN, JUDICIAL ATTITUDES IN SENTENCING 29-63 (1961) (discussing effects of legal and nonlegal factors); H. KALVEN & H. ZEISEL, *supra* note 6, at 115-16 (slightly more than half of judge-jury disagreement found to be explained by evidentiary factors, with other disagreement attributable to jury "sentiments on the law" or defendant; however, disagreement on evidence may occur because jury also "give[s] expression to values and sentiments under the guise of answering questions of fact"); Lempert, *Modeling Relevance*, 75 MICH. L. REV. 1021 (1977) (analytical exposition of effects on judicial decision-making of relevant and irrelevant information).

40. See J. Lachman, *supra* note 8, at 113-15 (statistical result associating evidence-related information and defendant characteristics with greater estimated strength of prosecutor in felony property cases in Detroit).

41. See H. KALVEN & H. ZEISEL, *supra* note 6, at 160 (table 52) (showing generally lower rates of jury acquittal when prosecution's evidence is stronger and when "balance of contradictions" favors prosecution); L. MATHER, PLEA BARGAINING OR TRIAL? 41-44 (1979) (discussing factors increasing defendant strength); J. Lachman, *supra* note 8, at 122-35, 149-52 (statistical results and scatter plots depicting bargain outcome and relative strength of parties).

42. The prosecutor can influence the development of a strong case by engaging in "sleuthing" activity, building upon the evidence initially provided by the police. The concept of relative strength employed in this Note refers to the level of strength prior to such explicit efforts on the part of the prosecutor. The latter type of action is reflected in the allocation of resources on the appropriate constraint.

Minimum Penalties

decision concerning the allocation of agency resources. With a budget of a given size and fixed cost commitments known, the remaining resources can be allocated among cases so as to achieve maximum agency satisfaction with the outcomes in the agency's portfolio of cases.⁴³ Let

43. This allocation decision is analogous to that of a consumer with a fixed budget making choices among alternative goods, as is illustrated in Figure B. Given the values

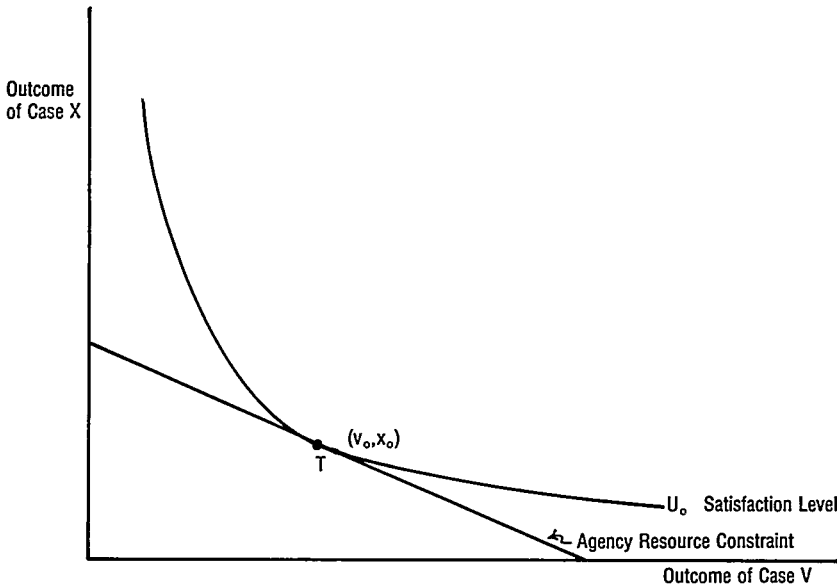


Figure B. Allocation of Agency Resources Between Cases V and X and Corresponding Trial Outcomes.

r_0 and M_0 of the parameters in the resource constraint, one can solve the constrained maximization problem and thus determine the values of v , x , and U associated with the optimal trial outcome. The equilibrium conditions are that the resource constraint be met and that the marginal rate of substitution of v and x be equal to r , that is,

$$\frac{\partial U(v,x)/\partial v}{\partial U(v,x)/\partial x} = r.$$

The result will be an outcome of v_0 for the instant case, and an outcome level of x_0 for the remainder of the portfolio.

An increased level of allocable resources would correspond to a northeasterly shift on the graph of the budget constraint, a more favorable initial strength level would be represented by a counterclockwise rotation of the constraint, and either of these would yield a higher outcome level in the instant case as well as, in general, in case X. The bargaining cutoff, or the reservation settlement, will generally rise with the level of anticipated trial outcome, and would rise in either of the examples given above. See note 41 *supra* (likelihood of jury trial acquittal or of low-sanction plea bargain outcome declines as strength of prosecution's case increases).

The exception in consumer theory to the relationship derived here for $v = f(r)$ is that of a Giffen good for which a price rise results in a greater, rather than lesser, quantity demanded. In order for this effect to occur, the good must account for a relatively large share of the consumer's expenditures, see J. HENDERSON & R. QUANDT, *supra* note 22, at 34, and this is unlikely to be true for a single case in the enforcement agency's portfolio.

the enforcement agency's preferences be represented by the function $U = U(v, x)$,⁴⁴ where v denotes the outcome of case V , the instant case, and x represents a composite of the outcomes of all other cases in the docket,⁴⁵ case X . If the agency allocates resources in an optimal fashion, given the level of its allocable resources and its initial level of strength relative to that of the defendant, the result will be an outcome of v_0 for the instant case, and an outcome of x_0 for the remainder of the portfolio.

The Trial Opportunity Set. The enforcement agency or prosecutor thus faces an opportunity set that pairs outcomes attainable at trial for the particular case V with those attainable for all other cases X , depending upon how resources are allocated. This resource constraint is represented by the equation $rv + x = M$, where r represents the relative strength or initial position of the parties,⁴⁶ and M represents the resources that the agency can allocate between the two cases.⁴⁷

2. *The Negotiated Outcome*

If a case is resolved by pretrial settlement, the agency can avoid the cost of a full trial and some of the case-maintenance costs otherwise incurred in the pretrial period. The pretrial settlement resource constraint is then $r'v + x = M'$, where M' is the new level of agency dis-

44. It is assumed that $U(v, x)$ is continuous and differentiable over the domain, $v, x \geq 0$, and that the usual assumptions about diminishing marginal utility with respect to v and x apply.

45. It is assumed that the agency allocates resources among the remaining cases—that is, case X —in an optimal fashion.

46. The more general form of the budget constraint is $rv + kx = M$; for simplicity, k is assumed equal to unity.

47. Outcomes v and x may be measured in various ways, depending on the institutional setting. For example, v could represent a term of incarceration, measured in years, or it might be an index of several aspects of the outcome, such as the weighted sum of a fine and a prison sentence.

The resource constraints need not be linear, but linear forms are used here for ease of exposition. The more general functional forms yield similar results. In the general case, the opportunity set is a production possibility frontier derived from the production functions for case development stated in terms of the relevant inputs, such as attorney time, and a budget constraint for the agency, stated in terms of the prices of the inputs to case development and the level of financial resources over which the agency exercises allocative control. J. Lachman, *supra* note 8, at 29-34. With the usual assumptions of diminishing marginal productivity, the resulting production possibility frontier corresponding to the opportunity set used in this Note has a negative second derivative for $v, x \geq 0$.

In addition, it is assumed that the relationships of the model are deterministic, although other approaches are possible. See, e.g., Becker, DeGroot, & Marschak, *Stochastic Models of Choice Behavior*, 8 BEHAVIORAL SCI. 41 (1963). For simplicity and tractability, the model employed in this Note assumes deterministic, rather than stochastic, relationships.

Minimum Penalties

posable resources, where $M' > M$,⁴⁸ and where r' is the relative strength of the parties under conditions of pretrial settlement. The value of the v -intercept of the new opportunity set is lower than before, because the best possible outcome from the agency's vantage point is less favorable than the best outcome obtainable from trial. Given this opportunity set, the agency is able to determine an optimal outcome pair (v', x') satisfying the new settlement opportunity set. If the corresponding utility level exceeds that obtainable from trial, the agency will be better off accepting a settlement.

If the new utility U' obtainable from settlement does not exceed that available from trial, however, the agency can still find settlement attractive if the defense offers a sufficiently large initial concession.⁴⁹ In particular, the defense can promise to refrain from using the full strength of its evidence when the settlement agreement is entered in court, thus confronting the agency with an artificially different value, r^* , of the strength parameter in the settlement opportunity set.⁵⁰ With sufficient, but customary, restrictions on the relevant functions,⁵¹ there exists exactly one level of strength-concession by the defense associated with settlement strength level r^* , and one level v^* , the minimally acceptable settlement, that would render the agency indifferent between the two modes of resolution.

The prebargaining stance of the agency is then to accept a settlement only if $v \geq v^*$. In deriving v^* and r^* , the equilibrium conditions are (1) that the satisfaction or utility level attained by pretrial settlement be equal to that attained from trial, U_0 ; (2) that r^* be equal to the marginal rate of substitution at that utility level, U_0 ; that is, in graphical terms, that the settlement opportunity set be tangent

48. The level of agency allocable resources M' is greater than before because of the cost saving from avoiding trial. Within this framework, the decision to drop a case may be viewed as a limiting case of pretrial settlement, in which the outcome is zero.

49. Prosecutors describe such an initial concession as finding out "if we're really talking about a plea." J. Lachman, *supra* note 8, at 57 n.34 (from interview with prosecutor); cf. A. DOUGLAS, *INDUSTRIAL PEACEMAKING* 278 (1962) (labor mediator cautions management to make sufficiently large initial concession in order not to insult labor representatives or damage negotiations beyond repair).

50. Such withholding of the full strength of a party's case is frequent in the handling of guilty pleas. Each party submits just enough evidence to suggest that the new charge, or suggested sentence, is appropriate to the case. D. NEWMAN, *supra* note 8, at 15-20. Actually, as Newman points out, the condition of appropriateness is not always met. *Id.* at 100-01 (citing examples); see, e.g., Remington, *Fair and Certain Punishment* (Book Review), 29 VAND. L. REV. 1309, 1315 (1976) (relating example from interview with inmate: "I see you were convicted of unarmed robbery in Detroit. What caliber of gun did you use? Without even a smile, the inmate would respond 'A .38 caliber revolver.'")

51. These restrictions are that the utility function and the trial and settlement opportunity sets meet the usual assumptions as described above. See notes 45-47 *supra*.

to the indifference curve corresponding to satisfaction level U_0 ; and (3) that the settlement opportunity set, $r' + x = M'$, be satisfied for $r' = r^*$.⁵² The determination of the settlement outcome is illustrated in Figure 1.

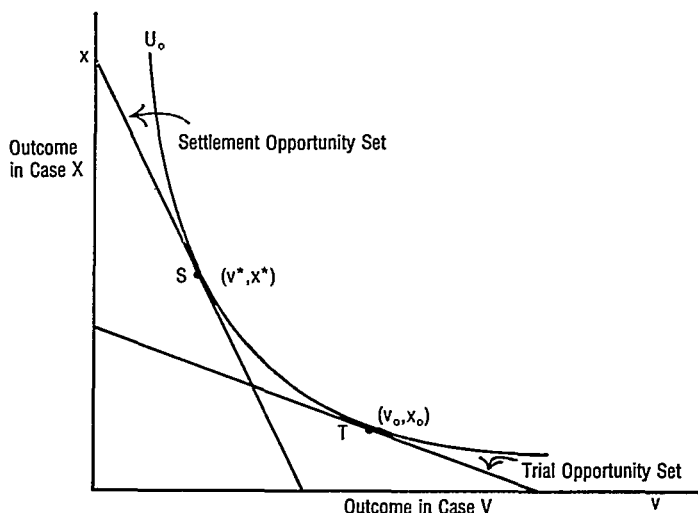


Figure 1. Equally Attractive Outcomes by Trial and Settlement

Since v^* was determined for an initial value of r , we can associate v^* with that value and label $v^*(r)$ the reservation settlement function, which describes the conditions under which the agency choice switches from trial to pretrial settlement. Permitting r to assume alternative values and v^* to adjust accordingly generates this reservation settlement function, which states the agency's demand for pretrial settlement in terms of the "price" of trial. As the defense's strength increases relative to that of the agency—that is, as r increases—the penalty that the agency is prepared to accept in settlement diminishes: pretrial settlement is a decreasing function of r .⁵³

52. The equilibrium conditions are:

- (1) $U(v^*, x) = U_0$,
- (2) $r^* = - \left(\frac{dx}{dv} \right)_{U_0}$, and
- (3) $r^* v^* + x = M'$,

where r^* is the level of strength exerted by the defendant.

53. That is, $\partial v^* / \partial r < 0$. Cf. Forst & Brosi, *supra* note 32, at 179-84 (alternative functional forms, also continuous, associating probability of conviction with case outcome); Landes, *supra* note 8, at 64-66 (same); Rhodes, *supra* note 35, at 315 (same).

Minimum Penalties

B. *Effects of Introducing Minimum or All-or-Nothing Penalties*

Minimum or all-or-nothing penalties may significantly affect the resolution of cases by trial or negotiation. They influence bargaining by altering the anticipated trial outcomes, which in turn modify the agency's preferred settlement outcome.

1. *Trial Outcome Effects and Consequences for Settlement*

To determine its bargaining position, an agency must first determine its anticipated trial outcome under the minimum penalty restriction. Holding r and M constant, let v_m denote the required minimum sentence, and let v_t be the lowest level of unrestricted trial outcome for which the decisionmaker remains willing to find liability or guilt, given the imposition of a mandated minimum. The relation of v_m and v_t is expressed by the equation $v_t = f(v_m)$, where v_t increases with v_m . This conviction threshold, v_t , need not equal the required minimum sentence; indeed, the extent to which these two values differ provides one measure of the degree of fulfillment of policy objectives. Two limiting cases of this relation are $v_t = v_m$ and $v_t = 0$. In the first case, decisionmakers nullify in all cases having an unrestricted trial outcome lower than the minimum penalty, and the policy fails.⁵⁴ In the second case, the legislative objective is fully im-

54. In less technical terms, the problem may be illustrated as follows. Suppose that two parties, defendants in separate actions, face only slight differences in evidence against them, but one defendant operates under the handicap of a legally irrelevant attribute. Without the imposition of a restriction on discretion, the first defendant would anticipate a fine of \$800 and the second of \$700, due to the difference in initial position. Once a minimum penalty of \$1000 is introduced, however, the decisionmaker may obey the policy in the former defendant's case, finding a \$1000 penalty preferable to no penalty at all, but may nullify in the latter case, feeling that increasing the penalty by almost 50% is going too far. By such a mechanism, a small piece of relevant evidence or a legally irrelevant factor, which previously resulted in a discernible but not dramatic difference in outcome, is associated with greater disparity in outcome for two cases that are otherwise identical. Cf. *Beck v. Alabama*, 447 U.S. 625, 642-43 (1980) ("apparently mandatory" death penalty "interject[s] irrelevant considerations into the fact finding process . . . [and] may encourage the jury to convict for an impermissible reason . . . [or] to acquit for an equally impermissible reason").

The bargaining effects are as follows: suppose that the negotiated outcomes in the absence of the penalty restriction would be \$400 and \$350, respectively. After introduction of the minimum penalty, however, the change in trial outcomes—from \$800 to \$1000 for the first case and \$700 to \$0 in the second—will tend to push the pretrial settlement outcomes farther apart than before. The first defendant will be seeking to avoid a \$1000 fine rather than one of \$800 and will be willing to pay more to do so; the second will be seeking to avoid an anticipated fine of \$0 and will be willing to pay less than when the alternative was a trial outcome of a \$700 fine. See p. 622 *infra* (Figure 4).

The effect of introducing the minimum penalty restriction for trial outcomes and for bargained outcomes benchmarked to them, is, therefore, to truncate what is otherwise a "smooth" relationship between overall initial strength level and the outcome of the dispute. See pp. 623-25 *infra*.

plemented: the mandatory sanction is assigned in all cases that under the unrestricted regime would result in any penalty at all.

In the agency's assessment⁵⁵ of the outcome of trial, the effects of introducing a restriction on penalty may be divided into four distinct situations: (1) the anticipated trial outcome in the unrestricted case exceeds the new penalty minimum, and the trial outcome is unaffected; (2) the anticipated trial outcome in the unrestricted case also exceeds the new required minimum, but the new minimum, even though not binding, nevertheless influences outcome; (3) the new minimum exceeds the anticipated trial outcome in the unrestricted case and raises the trial outcome to the new penalty level; and (4) the new minimum exceeds the anticipated trial outcome in the unrestricted case and results in nullification. These outcomes are illustrated in Figure 2 (on following page).⁵⁶

55. Analysis of the effects on the defendant's decisions of introducing a minimum penalty yields analogous results. There exists a value r_d of prosecutor strength relative to that of the defendant, below which the nullification prospect results in a reservation price more favorable to the defendant than would hold in the absence of the minimum penalty. A theory of bargained outcome that involves a proportional split of the bargainable difference yields a result for bargained outcomes that follows the pattern described in the text for the stance of the prosecutor. See J. Lachman, *supra* note 8, at 84-98, 135-36.

56. Depending on the form of the preference function, a fifth case may arise in which the optimal outcome in the unrestricted case lies in the region below the nullification threshold, but is nevertheless surpassed in desirability by the outcome combination for v and x associated with the application of the minimum penalty. This situation is illustrated in Figure C below. In this case, the introduction of the penalty requirement results in a decreased level of outcome in case X; that is, a reallocation of agency resources away from the rest of the caseload in favor of improving outcome in case V.

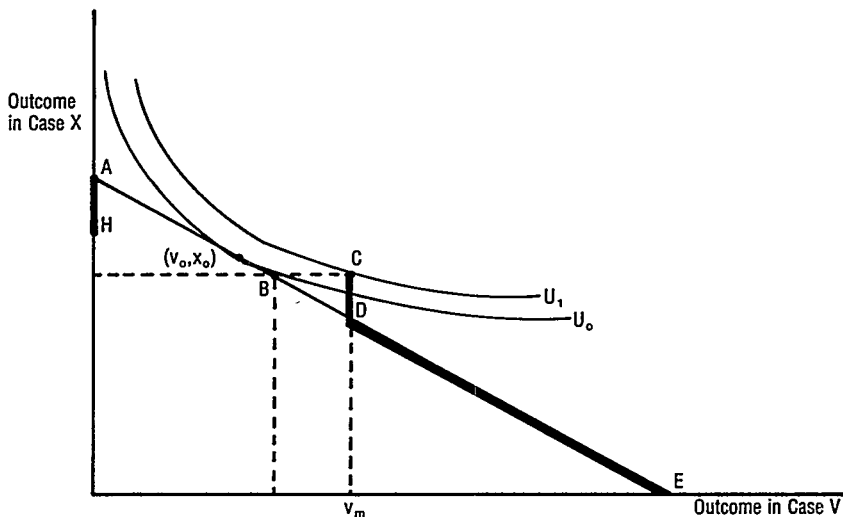


Figure C. Fifth Case, in Which Agency Allocation Shifts Toward Attainment of Minimum Penalty in Case V, with Consequent Allocation Away from Case X.

Minimum Penalties

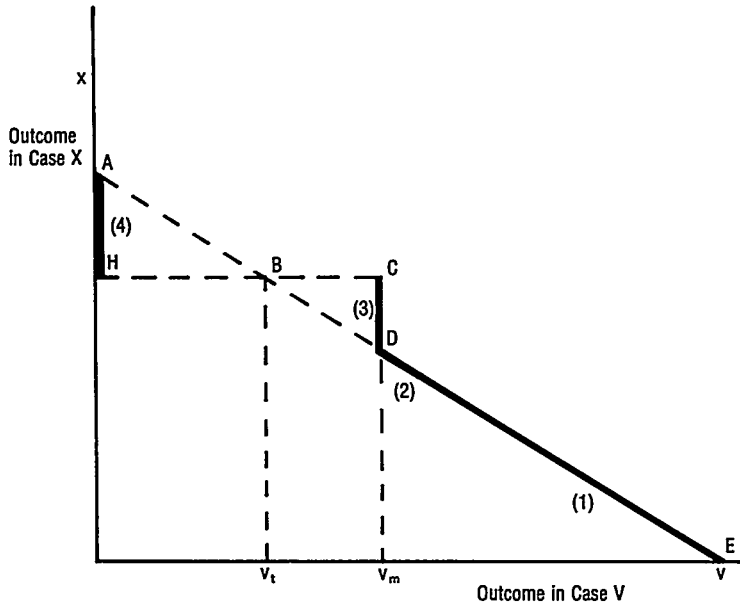


Figure 2. Four Cases of Anticipated Trial Outcome: (1) trial outcome unaffected by penalty requirement; (2) minimum penalty is not binding, but influences outcome; (3) policy raises outcome to new minimum; and (4) nullification.

The first situation, in which trial outcome remains the same as in the unrestricted case, results in the same bargained outcome as before. The second situation, although perhaps not common in practice, is more complicated. Assuming for the moment that $v_t < v_m$, then even if the unrestricted outcome is greater than the minimum penalty, v_m , it is possible that the minimum penalty will influence the outcome, and do so in a surprising manner. In this situation, the minimum penalty provides a comfortable fallback position for the agency in cases that otherwise afford an outcome above, but only slightly above, the new minimum.

The agency can obtain the same trial outcome in this second case—namely, the minimum penalty—even if the allocation of its resources results in a lower expenditure on case V than would occur in the absence of the new policy; this is true provided that the expenditure level does not fall below the level corresponding to that of outcome v_t . Because of this ability, the agency is able to reach a higher level of satisfaction by spending just enough to attain trial outcome v_t and by devoting the resources thus saved to the other cases in its portfolio. In this situation, the policy is successful in that it helps the agency,

even though it causes trial outcome to go down rather than up.⁵⁷ Despite this fall in trial outcome, the agency will require a higher reservation settlement level, which must provide satisfaction equal to the new higher level now attainable from trial.⁵⁸

In the third case, each anticipated trial outcome between v_t and v_m is mapped to v_m . This set of trial outcomes may be called the policy implementation zone. Within this zone, the utility-maximizing agency allocates as few resources to outcome improvement as possible, so long as the outcome is at least at level v_t , thus avoiding nullification.⁵⁹ In this situation, the agency obtains a more favorable outcome in case V than before and is still able to reallocate some of the resources saved to improving the outcome in other cases. As a consequence, the agency obtains a higher level of satisfaction from the prospective trial outcome and is correspondingly more demanding in terms of pretrial settlement.

The fourth, and final, situation involves the possibility of nullification, an outcome that puts the agency at a disadvantage. The restrictions on judicial discretion result in zero outcome for those cases with anticipated trial outcomes less than v_t ; that is, for cases that fall in the nullification zone. For the limiting case of full nullification, that is, for $v_t = v_m$, the higher the level set for the minimum penalty, the lower the level of agency satisfaction,⁶⁰ and the lower the level

57. Outcome v_m can be obtained at a resource commitment of only that required for v_t ; although the outcome of case V falls, the agency is able to reach a higher level of satisfaction overall, and is therefore aided by the policy. In contrast, in the nullification case, a lower trial outcome signifies policy failure; therefore, this second case is a distinctly pro-agency anomaly.

58. This circumstance, in which the lower trial outcome is associated with the higher reservation settlement level, is the exception to the general benchmarking effect of trial and settlement outcomes. The benchmarking effect is violated here because the reservation settlement level follows the change in *utility* associated with trial outcome, and the trial outcome in this special case does not reflect the overall well-being of the agency's outcomes on the whole portfolio of cases. Indeed, whenever the presence of the minimum penalty allows a higher level of utility to be reached, the reservation settlement is made higher by the minimum penalty.

59. As discussed above, *see* note 56 *supra*, there is also a special case in which anticipated trial outcome is below v_t , but the enforcement agency would choose to allocate resources toward improvement of outcome in case V in order to benefit from the "bootstrap" effect of the minimum penalty, as described in the text. In other words, the agency would obtain maximum benefit for this special case and for the one described in the text by trying to achieve *exactly* outcome v_t , whether by adjusting expenditures upward or downward in order to do so. The agency would allocate to case V just enough resources to avoid going into the zone of the fourth situation, where nullification is anticipated.

60. From the analysis of the preceding section and the assumptions distinguishing this case, it follows that there exists a point (v_o, x_o) in the original trial opportunity set

Minimum Penalties

of the reservation settlement. This result occurs because as the minimum penalty increases, it forces greater and greater reallocations of agency resources toward the improvement of outcome in case V, to the detriment of outcome in case X. Since the original allocation yielding outcome (v_o, x_o) would be optimal in the absence of restrictions on discretion, the effect of imposing ever higher levels of minimum penalty is to drive the best attainable allocation ever farther from the optimum. The successively worsening position of the agency with respect to trial outcome is reflected in a worsening position in pretrial negotiations.

that is associated with the optimal allocation of agency resources in the event of trial, and a corresponding point (v^*, x^*) denoting an equally satisfactory outcome if case V is resolved by pretrial settlement. The v^* value is the reservation price corresponding to trial outcome v_o . The minimum penalty restriction on outcomes from trial is associated with cutoff point v_t , where $v_t > v_o$ in this event. Since (v_o, x_o) represents an optimum position, with $v_t = v_m$, the point $(v_t, F(v_t))$ must necessarily be associated with a lesser level of satisfaction, U^* . Permitting adjustment through the settlement opportunity set, as before, one can determine the values of v' , x' , and r' such that the satisfaction gained

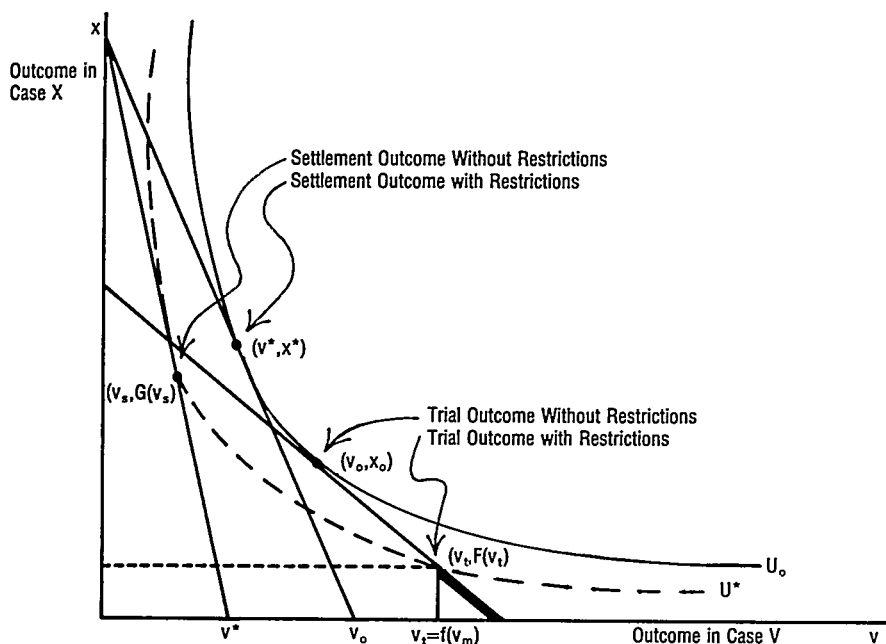


Figure D. Case Outcomes With and Without Restrictions on Judicial Decision.

from settlement at some point $(v_s, G(v_s))$ on the settlement opportunity set is exactly equal to that from outcome $(v_t, F(v_t))$ obtained at trial. The relationships of the trial outcomes with and without judicial restrictions to their settlement counterparts are illustrated in Figure D.

2. Effects on Outcome of Differences in Relative Strength of the Parties, Agency Resources, and Minimum Penalty Levels

In the absence of restrictions on judicial discretion, higher levels of initial strength of the defendant relative to the enforcement agency are associated with decreased trial and settlement outcomes.⁶¹ This relationship is represented by the continuous functions denoted $v(r)$ and $v^*(r)$, respectively. When a minimum penalty requirement is introduced, however, the trial opportunity set becomes the discontinuous relation depicted in Figure 2. Determining the effect that alternative levels of initial strength have on trial and bargained outcomes is more laborious than in the unrestricted case.

Let $F(r_1)$, shown in Figure 3 by the solid line, represent the trial opportunity set associated with relative strength level r_1 of the defendant, as in Figure 2. For $r_2 > r_1$, the set $F(r_2)$, shown in Figure 3

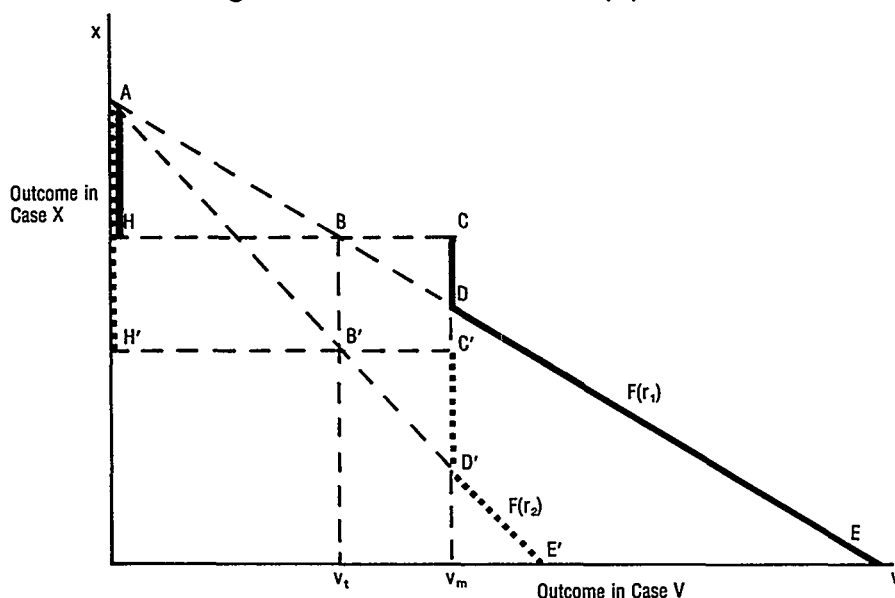


Figure 3. Opportunity Sets $F(r_1)$ and $F(r_2)$, $r_2 > r_1$.

by dotted lines, is determined as follows. First, the nonbinding portion of $F(r_2)$, segment $D'E'$, is smaller than that of $F(r_1)$, segment DE , with the result that a larger part of the opportunity set falls into the nullification and policy implementation zones. Second, for v in the implementation zone, that is, for $v_t \leq v \leq v_m$, the set of unrestricted

61. The exception, as before, is the unlikely case of a Giffen good. See note 43 *supra*.

Minimum Penalties

outcomes of segment $B'D'$ is mapped to segment $C'D'$, the latter being greater in length than its counterpart segment CD on $F(r_1)$.⁶² Third, the mapping to zero of the unrestricted outcomes for which $0 \leq v < v_t$ yields the nullification segment AH' .⁶³ Finally, since segment $C'D'$ is greater than segment CD , the zone of the comfortable fallback effect increases with the strength of the defendant.

Comparing $F(r_2)$ with $F(r_1)$, it is clear that the stronger the defendant's position relative to that of the enforcement agency, the more restrictive the penalty's effect. This results in greater relative importance of both the nullification and policy implementation zones.⁶⁴ The resulting trial outcome, considered as a function of r , is discontinuous. For low values of r —that is, when the defense is weak—the outcome coincides with that of the unrestricted case; for an interval of r -values, the outcome equals v_m ; and finally, for sufficiently large values of r , the outcome equals zero. The location of these discontinuities depends on the specific form of the preference function.

For r low enough that nullification does not occur—that is, for $r < r(v_t)$, as depicted in Figure 4—the agency is always at least as well off with the restriction as without it, and its reservation settlement will reflect that fact: the reservation settlement will be at least as great as in the unrestricted case for those outcomes.⁶⁵ For values of r at which trial nullification is anticipated, however, the reservation settlement is lower than in the unrestricted case.

62. The proof that $C'D' > CD$ is an exercise in the geometry of similar and right triangles.

63. As has been true throughout the discussion of this section, the text result assumes that $v_a = v_t$; that is, it assumes that the thresholds for nullification and for agency resource allocation are coincident. If this is not the case—that is, if $v_a \neq v_t$ —then the text analysis holds, but with the term v_t replaced by $\min\{v_a, v_t\}$.

64. This result is consistent with the contention that the decisionmaker's definition of "beyond a reasonable doubt" becomes more restrictive as the associated penalty rises. Although such a redefinition of the doubt threshold for conviction could also be modeled in terms of an increased threshold for nullification, the qualitative result would be the same as that described in the text. Further exploration of the interaction between nullification thresholds of judicial decisionmaking and relative strength of the parties is outside the scope of this Note. See J. Lachman, *Juror Decisionmaking Under Conditions of Restricted Choice* (1980) (unpublished manuscript) (on file with *Yale Law Journal*).

65. The proof of this is an exercise in the geometry of similar and right triangles. The settlement function $v^*(r)$ will have a similar shape, although the number of points of discontinuity may be greater. There are additional permutations possible for the boundary points of the agency's trial nullification and implementation borderline with those emerging from the application of the reservation settlement algorithm to the minimum penalty problem. These additional permutations would add more points of discontinuity to the relationships depicted in Figure 5, but would not otherwise significantly alter the results.

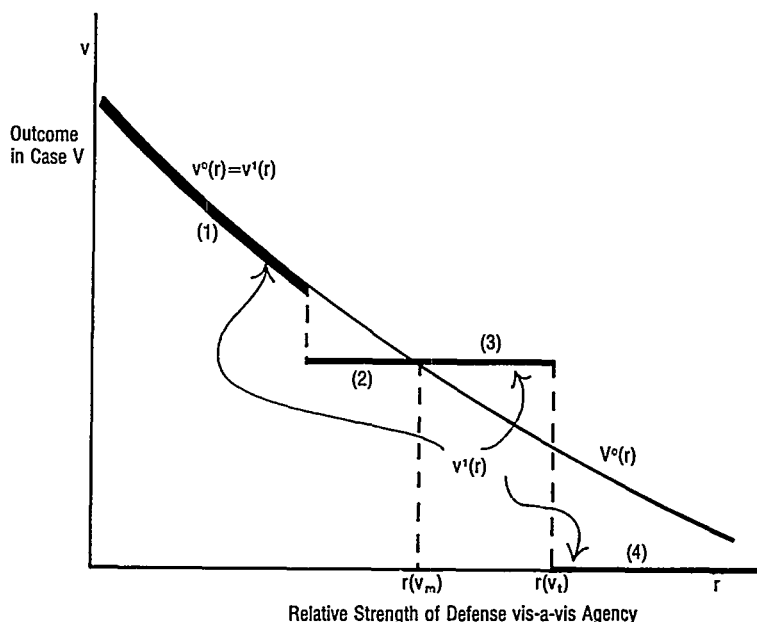


Figure 4. Trial Outcome as a Function of Relative Strength: without restriction $v^0(r)$; and with restriction $v^1(r)$ on trial outcomes. Numbers in parentheses correspond to those in Figure 2.

Increased levels of agency resources, M , have effects similar to those of a stronger agency initial position, as illustrated in Figure 5. A higher level of resources, as represented by $F^1(r)$, results in a larger range of outcomes for which the restriction is not binding.⁶⁶

The changes associated with an alternative level of the minimum penalty are also similar. A higher penalty shifts the level of v_m to the right, and increases the level of v_t , where $v_t = f(v_m)$. As a result, the nonbinding range is decreased in length, and both the nullification and the policy implementation zones are more prominent.⁶⁷ As the mandated penalty rises, the best obtainable trial outcome moves farther away from the unrestricted optimum, and the reservation settlement falls accordingly.⁶⁸

66. See pp. 615-16 *supra*.

67. A higher penalty is represented by a shift in the line $v = v_m$ to the right. Again comparing similar triangles and parallel lines, one obtains the result stated in the text.

68. This analysis does not predict the effect of agency effort on the remainder of the caseload without additional information or assumptions concerning the agency's preference function and constraints. In particular, under one set of circumstances, the introduction of the judicial restrictions will result in a shift of agency effort toward improving outcome among other cases in the portfolio; under other conditions, the remaining cases will suffer as well.

Minimum Penalties

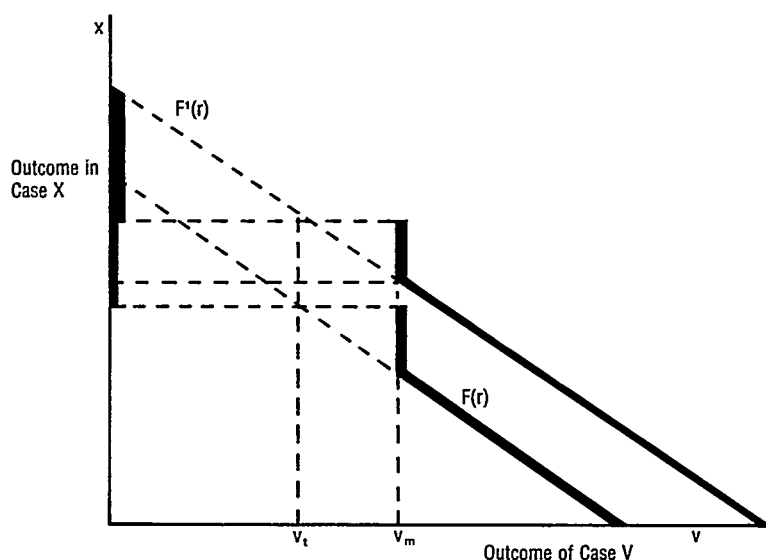


Figure 5. Effect of Increase in Resources from Level $F(r)$ to $F^1(r)$

III. Policy Consequences of Imposing Minimum Penalties

The introduction of an all-or-nothing or minimum penalty has several consequences that are of special policy interest. These consequences point to the need to consider not only minimum or all-or-nothing penalties, but also other available policy tools. Such an approach should attempt to minimize the divergence between penalties that satisfy the policy intent and those that satisfy the decision-maker's sense of fairness.

A. *The Effects of Minimum Penalties*

Consequences of minimum penalties include the effects on individual case outcomes and the implications of these changed outcomes for agency resource allocation and for the certainty and severity of sanction.

1. *Effects on Individual Outcomes*

When discretion is unrestricted, small increases in the initial strength level are associated with small changes in the anticipated trial outcome and in the reservation settlement level. When discretion is restricted by the imposition of a minimum penalty, however, the effects of differences in initial strength may change.⁶⁹ The adoption of a minimum

69. See pp. 620-22 & notes 61-66 *supra*.

penalty necessarily results in the exclusion from the set of possible trial outcomes of all levels of sanction between zero and the minimum penalty. The consequences for the defendant of losing the case obviously become more severe than in the absence of the penalty restriction. This change also produces important bargaining effects, which are similar to the effects on trial outcome.

With the imposition of a minimum penalty requirement, defendants with initial strength levels sufficiently great to yield a zero outcome—that is, acquittal—are clearly treated alike and, as a result, small differences in initial strength among members of this group do not become manifested in differences in outcome. Similarly, among those defendants assigned the minimum penalty, variation in initial strength levels among group members is not associated with variation in sanctions assigned.⁷⁰ This lack of change in penalty despite differing levels of initial strength may be desirable, for example, if the differences in initial strength are attributable solely to irrelevant information,⁷¹ or if the intent of policymakers is to effect uniform treatment of individuals whose cases exhibit only small differences in levels of initial strength. For two cases falling on the same side of the threshold for conviction, then, the introduction of the minimum penalty annihilates the effects of differences in initial strength of the parties.⁷²

When two cases fall on opposite sides of the threshold, however, the difference between them in initial strength is associated with a difference in outcome equal to the difference between the mandated penalty and zero.⁷³ This difference is greater than would have occurred in the absence of the minimum penalty.⁷⁴ Consequently, if the

70. See T. SELLIN & M. WOLFGANG, *THE MEASUREMENT OF DELINQUENCY* 45-50, 236-349 (1964) (interpretation of penalty structure as reflecting seriousness of offenses); A. VON HIRSCH, *DOING JUSTICE* 77-83, 132-40 (1976) (normative discussion of structure of penalties); Becker, *An Economic Approach to Crime and Punishment*, 76 J. POLITICAL ECON. 169, 177 (1968) (characterizing probability and severity of penalty as constituting a "price" of criminal behavior).

71. For the effect on bargained outcomes of varying levels of information, see L. FOURAKER & S. SEIGEL, *BARGAINING BEHAVIOR* 125-66 (1963) (bargaining experiments), and Harnett & Hamner, *The Value of Information in Bargaining*, 11 W. ECON. J. 81 (1973) (economic analysis).

72. As illustrated in Figure 4, two defendants having small differences in initial position and assigned penalties in the unrestricted case, reflecting the initial strength difference, may both be assigned the minimum penalty or both be assigned an outcome of zero.

73. This would occur, for example, if one defendant's case falls below and the other's above the nullification threshold v_i , as depicted in Figures 2 and 5. This jump corresponds to the borderline between the policy implementation and nullification zones, as defined above, see pp. 615-16 *supra*, and depicted in Figures 2 and 4. For discussion of indicators of success, see note 2 *supra*.

74. If the policy is fully implemented—that is, if there is no nullification—and if the penalty is applied in all cases yielding convictions in the absence of restrictions, then

Minimum Penalties

effect of a small difference in initial strength is to place one defendant on the acquittal side and another on the conviction side of the threshold, the difference in outcome associated with the small difference in initial strength will be magnified.⁷⁵ Since the initial strength of the parties was defined to be a function of various informational items, alternative levels of any one of these items, whether an item of admissible evidence or of irrelevant information,⁷⁶ can determine which outcome—acquittal or the minimum penalty—results.

2. *Effects on Agency Resource Allocation*

For the agency, the elimination of intermediate penalties and the resulting prospect of nullification make inadvisable some otherwise reasonable allocations of resources. In particular, an allocation that would yield an outcome substantially below that of the new minimum,⁷⁷ and consequently trigger nullification at trial and the bargain effects noted earlier, would waste resources that an enforcement agency could better conserve for other cases. This is true for all allocations yielding a trial outcome below the nullification threshold.⁷⁸

Thus, in the presence of such a policy, the enforcement agency is confronted with a new choice in cases in which a finding of liability would occur in the absence of a restriction on discretion. To make the best use of resources, it must either allocate resources heavily to the improvement of the instant case or reallocate its resources toward cases holding more promise for findings of liability through trial or settlement.⁷⁹

the threshold for conviction occurs at the lowest point at which any penalty would be applied at all. If the policy is not fully implemented, the extent of nullification determines the level of would-be outcome, in the unrestricted case, at which the threshold is located. This example illustrates the problem dealt with in more technical detail above. See pp. 615-16 *supra*.

75. This would occur, for example, if the absence of a single minor piece of evidence results in substantial numbers of acquittals and low settlements when intermediate levels of trial outcomes or bargained settlements would otherwise obtain. See Vidmar, *supra* note 6, at 215.

76. Such a problem would arise, for example, if jurors perceive the mandated penalty to be too harsh for whites, but not for blacks, as may be the case with mandatory death penalties. See note 54 *supra*; note 89 *infra*.

77. Determining the point at which this threshold is reached requires experimental research in which the level of the minimum penalty is changed by small intervals in successive experiments; the two-, three-, and four-verdict options of the Vidmar experiments, Vidmar, *supra* note 6, and the Lartz reanalysis of the Vidmar data, Lartz, *supra* note 24, do not provide enough options to approximate the effects of such small changes.

78. See pp. 616-19 *supra*.

79. We expect the agency to allocate additional resources among cases in such a way that the marginal benefit of an additional unit of resources is equalized for all cases in the portfolio; this is the usual requirement for economic efficiency, but with the addi-

This polarized pattern of adjustment in expenditure can generate mixed efficiency results. For those cases in which the policy is implemented and those in which the minimum penalty provides a comfortable fallback position, the new allocation of agency resources yields a higher level of agency satisfaction than would occur in the absence of the minimum penalty.⁸⁰ When nullification is anticipated, however, the resulting allocation is one that would not have been chosen in the absence of restrictions and is consequently associated with an efficiency loss,⁸¹ as viewed from the agency's perspective.⁸²

Moreover, the level of resources of a party and the costs associated with delay and duration of trial are closely intertwined. Obviously, the lower the costs associated with waiting for trial or carrying out the trial, the greater the resources left available for activities that improve trial and bargained outcomes.⁸³ When greater delay and its

tional complications of the indivisibility associated with the minimum penalty restriction.

This concerted reallocation of effort at a very early stage is exactly the sort of early screening recently begun in several urban jurisdictions with respect to criminal cases, and is undertaken in regulatory agencies as well. *See, e.g.,* LOS ANGELES FELONY PROSECUTIONS, *supra* note 25, at ix, 3 (discussing police and prosecution screening practices); R. NIMMER, DIVERSION: THE SEARCH FOR ALTERNATIVE FORMS OF PROSECUTION 14 (1974) (discussing "screening" out of system as possible official response in cases "calling for extreme leniency"); 99% of *Felony Arrests in the City Fail to Bring Terms in State Prison*, N.Y. Times, Jan. 4, 1981, § 1, at 1, 28, cols. 4-5 (estimate that majority of felony arrests were dismissed or reclassified as misdemeanors in boroughs where cases are screened by special staff units) [hereinafter cited as 99% of *Felony Arrests*].

Whether a case becomes a target of such agency effort depends in part on the initial strengths of the parties. Consequently, it is another potential magnifier of the effects of small differences in evidence or in legally irrelevant information. *See* pp. 610, 620-22 *supra*; notes 41 & 54 *supra*.

80. There are three cases in which the introduction of the penalty affords the agency a higher level of satisfaction by increasing the feasible set. The obvious case is the one of policy fulfillment; in addition, the second case noted in Figure 2, in which the new penalty provides a comfortable fallback position, and the fifth case, *see* note 56 *supra*, resulting in agency allocation toward case V, yield the same outcome.

81. In other words, the exclusion of a choice—in this case, of all those choices below the minimum penalty—results in the outcome being, at best, as good as the optimal one without the exclusion. This is the general reasoning in the construction of several price index measures. *See* J. HIRSHLEIFER, PRICE THEORY AND APPLICATIONS 140-45 (1976).

82. The unhappy alternative to this stringent early screening is to run the risk that resources would be devoted to cases that almost, but not quite, surpass the conviction threshold. Some losses of the latter kind may occur anyway because those who gather the evidence contributing to the prosecutor's initial strength, such as police officers, may not know in advance the strategy to be pursued by the enforcement agency in light of its allocable resources. *See, e.g.,* LOS ANGELES FELONY PROSECUTIONS, *supra* note 25, at ix, 62 (for over half of defendants arrested by police, District Attorney refused to file felony charges; of felony complaints filed, roughly half were rejected for felony prosecution).

83. *See, e.g.,* Japha & Bauman, *The Effects of the 1973 Drug Laws on the New York State Courts*, in STAFF WORKING PAPERS OF THE DRUG LAW EVALUATION PROJECT 128, 140 (Joint Committee on New York Drug Law Evaluation ed. 1978) (after introduction of minimum penalties for various drug offenses, "even cases which did not ultimately result in a trial took significantly more court time"); 99% of *Felony Arrests*, *supra* note 79, at 1, 28, col. 4 (Queens District Attorney cites backlog of "crisis proportions" despite

Minimum Penalties

consequent delay-associated diminution in evidence⁸⁴ occur in the presence of a minimum penalty restriction,⁸⁵ one may expect more frequent or more extreme nullification at trial and bargained outcomes different from those that the same evidence would have brought in the unrestricted case. In addition, relatively greater importance may be associated with evidence and with nonevidentiary factors contributing to the initial position of the parties,⁸⁶ thus magnifying differences in individual outcomes and leading to efficiency losses associated with lower levels of resources.

3. *Effects on Severity and Certainty of Sanction*

In addition to the effects noted above, the overall pattern of trial and bargained outcomes⁸⁷ determines levels of certainty and severity⁸⁸

felony prosecution of only one-fourth of felony arrests made in previous year). Since increases in delay or trial outcome are borne by both parties to a dispute, the net effect of such cost increases on the parties is not immediately obvious. In terms of the level of resources allocable to outcome-improvement activities, both parties are worse off because of delay, and their comparative disadvantages will depend on many factors tied to the individual case. For example, delay will be more costly for defendants incarcerated before trial than for those who are not. *See* p. 609 & note 38 *supra*. The initial strength of the parties will be of relatively greater importance in determining case outcome when delay is great, and resources that might otherwise be devoted to improving trial outcome are instead used for case maintenance and other costs of delay. *See* J. Lachman, *supra* note 8, at 37-43 (economic representation of costs of trial and of pretrial delay).

Added to the "pure" resources effect is the "burden" effect applicable as a function of various possible burdens of proof. For example, because the prosecution in a criminal case must prove guilt beyond a reasonable doubt, a diminished level of resources due to increased trial delay may make the difference, in terms of the associated level of outcome improvement activity that is possible, between reaching and not reaching the required proof threshold. In addition, because evidence available to a party tends to decline over time—as witnesses die, move away, or forget, and as real evidence is or is not maintained—delay will take a relatively higher toll from a party facing the heavier or more difficult burden of proof. *See* Japha & Bauman, *supra*, at 170-71, 177-78; J. Lachman, *supra* note 8, at 113-15, 119 n.22, 120 nn.23-25; Special Narcotics Unit of the City of New York, Progress Report, Centralized and Decentralized Parts at V-21 to V-23 (1976). The relative effect generally favors the defendant. *See* J. Lachman, *supra* note 8, at 113-14 (statistical results indicating that prosecutor strength declines as a function of delay until trial). The effect of the enforcement agency's burden of proof may be negated, however, in the event of presumptions favoring the agency. *See generally* Jeffries & Stephan, *supra* note 12; Underwood, *The Thumb on the Scales of Justice: Burdens of Persuasion in Criminal Cases*, 86 YALE L.J. 1299 (1977).

84. The evidence losses to each side will vary, depending on the kind of evidence primarily relied upon. For example, a party heavily dependent on witness testimony and memory will be more vulnerable to delay effects than a party that relies on documents, which can presumably be preserved intact. *See* F. CANNAVALE, WITNESS COOPERATION pt. 1, at 12-17, pt. 2, at 2-3 (1976).

85. *See* Japha & Bauman, *supra* note 83, at 139 (new mandatory penalty drug law resulted in growing backlog of cases in New York City).

86. *See* pp. 624-25 *supra*; notes 19-21 *supra*.

87. This is true also for cases dropped or not pursued by the enforcement agency because of the changed pattern of priorities among cases consequent to the adoption of the required penalty. *See* pp. 625-26 *supra*.

88. Such a change is consistent with social science findings of a negative relationship

of sanction different from those prevailing in the absence of the policy. In particular, if there is substantial nullification, the increase in severity of the prescribed sanction will be accompanied by a fall in the likelihood of the sanction actually being assigned, so that the increase in severity is obtained at the expense of a fall in certainty.⁸⁹ If, as current research indicates, certainty is more important than severity in deterring the proscribed activity,⁹⁰ then the new policy

between the level of sanction and the likelihood of conviction, and the characterization of likelihood and severity of sanction as substitutes for one another in deterring proscribed activity. See, e.g., Gibbs, *Crime, Punishment and Deterrence*, 48 Soc. Sci. Q. 515 (1968). But see Antunes & Hunt, *The Impact of Certainty and Severity of Punishment on Levels of Crime in American States: An Extended Analysis*, 64 J. CRIM. L. & CRIMINOLOGY 486, 492-93 (1973) (if certainty of sanction is low, increased severity will have little deterrent effect); Ehrlich, *Participation in Illegitimate Activities: A Theoretical and Empirical Investigation*, 81 J. POLITICAL ECON. 521, 544-47, 552 (1973) (same result obtained by multiple regression analysis).

It should be noted, however, that certainty is more important than severity *within the relevant range* of the certainty-severity tradeoff. Clearly, outside this range there may exist sufficiently small differences in certainty and sufficiently large differences in severity to reverse the relative order of their effectiveness for deterrence. This consideration is often ignored in the literature. Cf. Chiricos & Waldo, *Punishment and Crime: An Examination of Some Empirical Evidence*, 18 Soc. PROB. 200, 200-10 (1970) (in data analysis not taking certainty and severity into account simultaneously, correlation of certainty with criminality varies over time and among offenses and there is no consistent support for deterrent effect of severity); Tittle, *Crime Rates and Legal Sanctions*, 16 Soc. PROB. 409, 409, 419, 422-23 (1969) (severity of punishment "has little consistent independent or additive effect," although total effect of certainty and severity is complex).

For example, with the imposition of the minimum penalty, those cases in which the minimum, or more than the minimum, is assigned will tend to weight the severity factor upward. But cf. pp. 617-18 *supra* (discussing conditions under which bargain outcome is more severe, even though trial outcome would not be at minimum penalty or above). Each case, however, in which a zero penalty is assigned—whether through judge or jury acquittal, judge dismissal, or prosecutor decision to drop charges—pushes the likelihood factor downward.

It is also possible for there to be an appearance of nullification independent of its occurrence, with the perceived low probability of sanction having an effect on behavior similar to that of actual nullification. See Tittle & Rowe, *Certainty of Arrest and Crime Rates: A Further Test of the Deterrence Hypothesis*, 52 Soc. FORCES 455, 457-61 (1974) (deterrence effect exists when enforcement exceeds 30% arrest clearance rate, but if enforcement is below this "tipping point," differences in enforcement have only negligible deterrent effect).

89. This has been contended in particular for the imposition of a mandatory death penalty. See Riedel, *supra* note 6, at 265, 282-83 (empirical study of death penalty); pp. 615-16 & note 54 *supra*. The use of the term "certainty" in the deterrence literature, where it connotes likelihood, such as that of arrest, prosecution, or sanction, is different from the use of the term elsewhere in economics, particularly in portfolio theory.

90. See Chambers, *Men Who Know They Are Watched: Some Benefits and Costs of Jailing for Nonpayment of Support*, 75 MICH. L. REV. 900, 912-13, 936-37 (1977) (deterrence strongest in counties in which likelihood of prosecution and of jailing lead potential offenders to perceive clear link between their own behavior and system that leads to punishment); note 88 *supra* (discussing relative deterrent effect of certainty and severity of sanction). See also Bailey & Smith, *Punishment: Its Severity and Certainty*, 63 J. CRIM. L.C. & P.S. 530, 536-38 (finding inverse relationship between severity and certainty of sanction); Polinsky & Shavell, *The Optimal Tradeoff between the Probability and Magnitude of Fines*, 69 AM. ECON. REV. 880 (1979) (investigating optimal level of fines and likelihood of sanction taking into account cost of enforcement).

Minimum Penalties

may increase, rather than decrease, the level of occurrence of the proscribed activity.

B. *Policy Implications*

It is when the minimum penalty is high or when the enforcement agency's initial strength or allocable resources are relatively low that the policy has its "bite," altering the outcome otherwise reached by trial or negotiation.⁹¹ Paradoxically, these are also the circumstances under which the possibility of policy frustration looms greatest. The most successful policies will be those that minimize the disparity between the legally correct outcome and the outcome consistent with the decisionmaker's sense of fairness. Yet, if the purpose of the policy is to change substantially the results of judicial decisionmaking, then the policies having the greatest chance of success—those closest to what judicial decisionmakers would choose anyway—are the ones that least fulfill policy intent even when successful.

Under a policy of mandatory minimum penalties, some defendants who would otherwise receive lower penalties obviously suffer losses because of the new policy. If such defendants are the targets of the policy,⁹² this result represents fulfillment of policy objectives. Yet other defendants who would be convicted under a regime of flexible sentencing may go free as a result of nullification by judges or other decisionmakers. The resulting penalty reductions, as well as the aggregate effect of a drop in deterrence due to decreased certainty of sanction, constitute losses from the policy that are borne by the public.

The pattern of policy fulfillment versus nullification depends not only on the level of the penalty, but also on the level of the financial commitment to enforcement agencies. If the prosecuting agency's budget is greater, nullification plays a less prominent role;⁹³ in that event, some case outcomes are higher not only because of the new penalty requirement, but also because increased prosecutorial effort results in an anticipated outcome high enough so that the minimum penalty level is surpassed. Similarly, a financial commitment to the institutions providing the initial information may place the prosecutor in a position strong enough so that the minimum penalty need not be relied upon for obtaining favorable results.⁹⁴ Taking the resource

91. See pp. 620-22 *supra*. For discussion of criteria of policy success, see note 2 *supra*.

92. The policy is successful in the sense used here if the affected group is neither underinclusive nor overinclusive as compared with the target group.

93. See p. 622 *supra*.

94. See pp. 620-24 *supra*. This Note does not address the question of the proper distribution of resources among various enforcement and legal process institutions; such resources should, however, be distributed in such a way that the marginal benefit of the

commitment into account, the policy decision involves the weighing of the goal of increased penalties in some cases against the prospect of nullification in others.

In addition, there are other factors to be considered, including the incidence of policy effect across individuals due to small differences in evidence or differences in legally irrelevant information. Under some circumstances, the policy results in like treatment of individuals who, under a regime of no restrictions, would be treated in a disparate fashion.⁹⁵ Under other circumstances the presence of the penalty serves to magnify the disparity of treatment accorded those who differ only in irrelevant ways or whose cases exhibit only a small difference in evidence.⁹⁶ This pattern of effect may or may not be socially desirable, but in either event it can lend substantial weight to one side of the balance and thus be critical to the policy choices made.

Finally, there are considerations not taken into account in the model presented here that should nevertheless be relevant to a final policy decision. Among these are judgments about the importance to the community of making a strong symbolic statement about community values—for example, by the adoption of a death penalty—despite possible nullification in practice. By using the law to express community preferences in the abstract while permitting judges and juries to exercise individualized judgment on a case-by-case basis, the community is able simultaneously to accommodate a commitment to uniformity of treatment and to satisfy a desire for fairness to individuals.⁹⁷

last dollar allocated is equalized for all institutions subject to the allocation decision. See Thurow, *Equity Versus Efficiency in Law Enforcement*, 18 PUB. POLICY 451 (1970) (discussing distribution of resources by marginal benefit rule, where definition of "benefit" varies as a function of equity goals).

95. See pp. 620-22 & notes 61-66 *supra*.

96. *Id.*

97. One approach to the problem is to adopt legislatively specified adjustments for aggravating and mitigating circumstances. See TWENTIETH CENTURY TASK FORCE ON CRIMINAL SENTENCING, *supra* note 25, at 19-34 (proposals); A. VON HIRSCH, *supra* note 70, at 79-106, 124-31 (report of Committee for the Study of Incarceration). The latter proposals, allowing for the adjustment of mandated sentences if aggravating or mitigating circumstances are shown, capture some but not much of the flexibility needed to obviate the undesirable effects of mandated penalties discussed in this Note. Cf. Underwood, *Law and the Crystal Ball: Predicting Behavior with Statistical Inference and Individualized Judgment*, 88 YALE L.J. 1408, 1426 n.49 (1979) (noting conflict between "fundamental value of decisionmaking in accordance with general standards and the equally fundamental value of individualized decisionmaking"). The adoption of uniform penalties is generally considered to be a step in favor of the general standards approach, but, as demonstrated in this Note, the effects may be to increase disparity in the treatment of individuals.

Minimum Penalties

In the end, the choice is not simply one of whether or not to adopt a penalty requirement, but a choice about a broader and more complex set of policy options, and, in effect, a choice among policy consequences. For a specified level and form of sanction, the extent of policy fulfillment, nullification, and other effects can be estimated, given sufficient research before adoption of such policies. In light of these anticipated consequences, decisions that are essentially political must still be made. Since the prospect of nullification looms most prominently at those levels of penalty for which the policy significantly changes judicial outcome, the extent of policy fulfillment and its associated benefits must be weighed against the losses associated with nullification and other policy consequences.

Ultimately, these choices are made not in the abstract, but in the context of specific legislative goals and values. This Note suggests that those policies will be most successful—both in terms of trial and bargained outcomes, and in terms of the influence on decision processes in the courts—that take into account the anticipated responses of judicial decisionmakers, and the institutional tradeoffs that their decisions precipitate.